

Metodología para la segmentación automática de la corteza cerebral sobre imágenes MRI basada en características volumétricas usando técnicas de renderizado tridimensional por funciones de transferencia

Trabajo de grado para optar al título de Magíster en Ingeniería Eléctrica

Estudiante: Jhon Jairo Castañeda Gonzalez

Director: Genaro Daza Santacoloma

Co-director: Álvaro Ángel Orozco Gutierrez



**Universidad Tecnológica de Pereira
Facultad de Ingenierías - Programa de Ingeniería Eléctrica
Maestría en Ingeniería Eléctrica
Grupo de Investigación en Automática
Pereira, Risaralda, Colombia
2019**

Contenido

I	Preliminares	4
1.	Introducción	6
2.	Objetivos	10
2.1.	Objetivo general	10
2.2.	Objetivos específicos	10
II	Marco Teórico	11
3.	Renderizado volumétrico	12
3.1.	Funciones de transferencia	13
3.1.1.	Función de transferencia Gradiente	14
3.1.2.	Función de transferencia Laplaciana	17
3.1.3.	Función de transferencia Valores LH	18
3.1.4.	Función de transferencia Curvatura	20
3.1.5.	Función de transferencia Estadística	22
3.1.6.	Función de transferencia Escala	22

4. Análisis de relevancia	25
4.1. Análisis de relevancia estocástico	25
4.2. Puntuación Laplaciana	26
4.3. Información mutua	27
4.3.1. Aproximación por entropía de Renyi	29
4.4. Estrategias de búsqueda secuencial	30
4.4.1. Algoritmo COBRA	31
 III Marco Experimental	 32
5. Esquema general de trabajo	33
5.1. Descripción de bases de datos	34
5.2. Procesamiento y caracterización	35
5.3. Selección de características	36
5.3.1. Selección mediante información mutua	36
5.3.2. Selección por métodos de puntuación	38
5.4. Clasificación	38
5.4.1. Clasificación de datos desbalanceados	38
 6. Resultados	 42
6.1. Procesamiento y caracterización	42
6.2. Análisis de relevancia	46
6.3. Clasificación	52

7. Conclusiones y trabajo futuro	56
7.1. Conclusiones	56
7.2. Trabajo futuro	57

Parte I

Preliminares

Resumen

El diagnóstico de anormalidades cerebrales es una tarea que plantea un reto para los especialistas, debido a que muchas de estas patologías requieren de métodos que resultan incómodos para los pacientes y en algunos casos no logran su cometido. La segmentación de estructuras cerebrales mediante imágenes médicas se plantea como una herramienta de apoyo eficaz y no invasiva para el apoyo al diagnóstico de neuro-patologías. Sin embargo, existen tres problemas latentes en la segmentación de estructuras, el primer problema es determinar un espacio de representación adecuado que puede revelar información de la anormalidad; el segundo se ubica en el desbalance de datos que es frecuente en estudios de tipo, que tiendo a hacer que las metodologías de segmentación fallen al detectar las anormalidades; y por último el problema de tener millones de datos como consecuencia de las imágenes médicas o datos volumétricos, lo cual hace que sistemas de computo no sean capaz de procesar la información. Por lo tanto, en este trabajo se plantea una metodología de segmentación de estructuras cerebrales que pueden manejar el desbalance y los altas cantidades de datos, encontrando un espacio de representación idóneo.

Capítulo 1

Introducción

El diagnóstico de patologías cerebrales se basa en la evaluación semiológica del paciente a través de manifestaciones motoras, funcionales y comportamentales que generan sospecha de algún tipo de déficit neurológico. Los estudios de Tomografía Axial Computarizada (TAC) o de Resonancia Magnética (IRM) permiten identificar y localizar anomalías anatómicas de manera no invasiva que confirmen el diagnóstico [1,2]. A pesar de esto, no todas las características morfológicas de las patologías son identificables a simple vista en un estudio imagenológico debido a que las características no guardan necesariamente relación con los fenómenos físicos en los cuales se fundamentan los sistemas de adquisición de imágenes médicas [3–5], o también se debe a que los equipos de TAC y IRM entregan solo cortes bi-dimensionales que no facilitan la percepción espacial de la morfología tridimensional de las estructuras anatómicas [6]. Por lo que, en ocasiones las patologías no son diagnosticadas o confirmadas a pesar de que se trate de un caso positivo [7]. Sin embargo, la identificación de algunas anomalías puede llegar a ser difícil si no es claro a qué corresponde un tejido en particular. Una de las lesiones de la corteza que más desafío presenta a los especialistas en su diagnóstico, son las displasias corticales focales (DCF), las cuales se definen como un trastorno en la organización neuronal y causa frecuente de la epilepsia refractaria. Estas son usualmente diagnosticadas mediante el estudio de imágenes de resonancia magnética (IRM). Sin embargo, este procedimiento es subjetivo y en muchas ocasiones se requiere de un grupo de expertos para su identificación. La identificación adecuada del tejido afectado garantiza un mejor resultado de reducción de crisis convulsivas cuando el tejido debe ser removido quirúrgicamente. Por tanto la precisión y la identificación de las zonas afectadas con DCF son una problemática importante en el tema médico. En la actualidad se reconocen una gran variedad de técnicas para facilitar a los especialistas la identificación de la zona afectada. Respecto a las DCF, estas representan un subconjunto de malformaciones del desarrollo cortical en el cual hay anomalías de la laminación cortical, maduración neuronal y diferenciación neuronal. Taylor *et al.*, [8] fueron los primeros en encontrar dicha patología, y describirla como una anomalía en la corteza

cerebral. Clínicamente las DCF implican deficiencias neurológicas, problemas cognitivos, y epilepsia, siendo la epilepsia la manifestación más común de las DCF, y aquella que más imposibilita la vida normal de las personas que la padecen [9].

En la actualidad, se dispone de una variedad de métodos de procesamiento de imágenes que facilitan, a los especialistas clínicos, la evaluación de las estructuras anatómicas. En general, los métodos empleados para la segmentación de tejidos se componen de una etapa de extracción de características y del mapeo de la información a un nuevo espacio de representación. En particular, para la identificación de anomalías en tejidos cerebrales a partir de imágenes de resonancia magnética se han empleado técnicas como la morfometría por voxel y del reformato curvilíneo, los cuales son métodos basados en la comparación de la IRM del paciente con modelos estadísticos construidos a partir de sujetos sanos [7, 10]. El mayor inconveniente de este tipo de técnicas es la necesidad de tener una base de datos amplia de sujetos sanos para la construcción de los modelos de referencia. Adicionalmente, la manera en que se realiza el diagnóstico depende en gran medida de la familiaridad del especialista clínico con el algoritmo, ya que los resultados se buscan manualmente sobre los mapas de características generados [4]. Por esta razón, no ha sido posible masificar el uso de esta metodología en la rutina hospitalaria [5].

La técnica comúnmente usada como “*Gold Standar*” (referente para comparar que tan eficaz es una nueva técnica) para la segmentación, está construida a partir de atlas cerebrales, construidos utilizando un gran número de sujetos de atributos similares (edad, género, medidas, entre otros). Sin embargo, para su uso existen varios inconvenientes que suelen conducir a errores en la tarea de segmentación. Se hace necesario que el atlas esté normalizado y adaptado a la imagen del sujeto de estudio, lo que en muchos casos puede implicar un gran reto, debido a la variedad anatómica del cerebro y a las condiciones en las que se toma la imagen médica. Adicionalmente, el especialista debe tener un conocimiento sólido de la fisiología del cerebro, pues la interpretación del atlas es totalmente subjetiva [4, 11].

Una herramienta que ha probado ser eficiente para el renderizado volumétrico, es la función de transferencia (FT) para visualización de volumen de imágenes médicas, la cual ha permitido la representación detallada de los datos contenidos en las imágenes. Es decir, permiten la visualización detallada de zonas de interés en el volumen, asignando a esos datos un color y una opacidad, haciendo que el resto de los datos del volumen sean invisibles mientras se visualiza la zona determinada, pero su uso no es intuitivo, lo que dificulta su aplicabilidad clínica [12]. Aunque las funciones de transferencia según una revisión parcial del estado del arte, no se han implementado para la detección de displasias, prometen ser una herramienta potencial para su ubicación. Las estructuras cerebrales, por ejemplo la corteza, tienen características superficiales y volumétricas, entre otras, que pueden ser calculadas a través de las FT. Para el caso de la corteza por ejemplo, la curvatura, la anchura y la profundidad de los surcos, son características que pueden ser adaptadas y calculadas por los espacios de FT, convirtiendo a estos espacios de FT en métodos afines para la segmentación y estudio de los

tejidos y estructuras cerebrales. Las FT como herramienta para el renderizado volumétrico, se basan en el modelamiento de la interacción de la luz sobre el medio. Cuando la luz atraviesa el volumen, esta puede ser emitida, absorbida o dispersada. Como consecuencia, se puede establecer una ecuación para representar la transferencia de luz. Sin embargo, dicha ecuación presenta bastante complejidad y su solución no es analítica, por ello, se presentan variantes simplificadas, que asignan características a los datos volumétricos [13, 14].

Los espacios de FT que se pueden generar, dado un proceso de cálculo, son en su mayoría espacios en los cuales es difícil reconocer un objeto o zona de interés (datos a estudiar) debido a que el usuario debe tener conocimiento de qué está buscando y cómo funciona el espacio de FT generado. Por otra parte, dado que el espacio no es representativo del objeto bajo estudio, la sintonización de la FT sobre este espacio (curva para los espacios 1D o Widget para los espacios de 2 o más dimensiones), se realiza por ensayo y error, careciendo de algún tipo de lógica y puede llegar a desaprovechar casi por completo la información ofrecida por el espacio donde vive la FT [15, 16]. Algunas investigaciones reportadas en el estado del arte proponen técnicas para sintonización automática de la FT, dado que la generación automática no requiere interacción con el usuario. Para la visualización de incertidumbre de superficies, Pfaffelmoser *et al.* [17] usaron una FT generada automáticamente para colorear superficies con base en la desviación espacial aproximada de los puntos de la superficie de la media [18]. Usando la separabilidad de celdas, aquellas celdas que potencialmente puedan contener información relevante se almacenan, mientras que aquellas que pueden ser producidas por ruido se descartan. Finalmente el usuario puede asignar valores de color y opacidad a una de las celdas o a un conjunto de ellas. En [19] y [20] han automatizado la asignación de color para las FT a partir de un punto inicial en 3D en un espacio topológico y la representación con grafos. Wang y Kaufman [21], han usado una función de importancia que el usuario puede modificar para calcular automáticamente la FT. Los objetos de importancia en el volumen son resaltados usando color. Bramon *et al.* [22] han usado estrategias basadas en teoría de la información, empleando conceptos como normatividad y divergencia informacional, para automáticamente definir FT multimodales, pero recayendo en el problema de intuición, puesto que necesitan conocimiento específico de lo que buscan en dichos espacios. Es decir, debe haber preparación técnica para su uso posterior. Por otro lado, muchas de estas FT necesitan experiencia para su aplicación, pues no cualquiera permite ver lo que el especialista desearía.

A pesar de la eficiencia que presentan los espacios de FTs para la tarea de segmentación e identificación de estructuras en datos volumétricos, los mejores resultados logrados que se reportan en el estado del arte, evidencian ser espacios de alta dimensión, lo que supone un reto de implementación [13, 14]. Debido a estos espacios de alta dimensión, la selección y la clasificación de características se presentan como tareas independientes, lo que puede suponer un problema al intentar obtener los mejores resultados según afirman [23–26], ya que se espera que el conjunto de características seleccionados explique casi en su totalidad el

conjunto completo de datos. Sin embargo, un conjunto de las mejores características, puede que no sea el conjunto con las mejores características, pues el no optimizar la selección de características para la tarea de clasificación, empeora su rendimiento [27, 28].

Capítulo 2

Objetivos

2.1. Objetivo general

Desarrollar una metodología de segmentación automática de la corteza cerebral sobre imágenes IRM con base en características volumétricas y renderizado tridimensional por funciones de transferencia, usando algoritmos de programación semi-definida para optimización de la segmentación, con medidas de información mutua, para generar una herramienta de apoyo al diagnóstico asistido de patologías asociadas con anormalidades de la corteza.

2.2. Objetivos específicos

1. Desarrollar una metodología de caracterización volumétrica de IRM utilizando espacios de funciones de transferencia, que permita codificar las propiedades estructurales como curvatura, homogeneidad, divergencia, entre otras, relacionadas con los diferentes tipos de tejidos cerebrales y facilitar la identificación de patrones discriminantes asociados con anormalidades de la corteza cerebral.
2. Desarrollar una estrategia de extracción de características supervisada basada en aprendizaje por teoría de información, en aras de construir un espacio de representación discriminante a partir de características volumétricas calculadas sobre IRM que permita diferenciar la corteza sana y enferma.
3. Desarrollar una metodología de segmentación automática de la corteza que incluya la estrategia de extracción de características propuesta, con el fin de identificar zonas de interés relacionadas con anormalidades de la corteza cerebral.

Parte II

Marco Teórico

Capítulo 3

Renderizado volumétrico

La exploración y recorte volumétrico juega un rol decisivo en el entendimiento de datos tridimensionales, permitiendo seccionar partes del volumen que puedan ser de interés, basados en las posiciones de los voxels en el set de datos. Esta tarea se ha convertido en una labor fundamental en aplicaciones de diagnóstico médico, debido a que la selección de las zonas de importancia es usualmente la única manera de resaltar detalles importantes en los datos que están ocluidos por otros [29].

El renderizado volumétrico es un conjunto de técnicas utilizadas para mostrar una proyección 2D en pantalla a partir de un conjunto de datos discretos 3D [30, 31], conservando las propiedades visuales de la estructura real. Las técnicas de renderizado pueden categorizarse según el método que utilicen, estos pueden ser algoritmos de «orden-imagen» o de «orden-objeto». Las técnicas de orden-imagen se inicializan en un voxel en particular y buscan en su vecindad aquellos voxels que compartan color y opacidad con el voxel objetivo. Mientras que las técnicas orden-objeto operan en un espacio objetivo (imagen 3D) donde los voxels son proyectados en una imagen 2D, y la información es extraída e interpretada según la tarea en cuestión. El renderizado volumétrico se basa en el modelo de interacción de la luz con el medio por el que pasa. Cuando la luz atraviesa un medio, puede ser emitida, absorbida y/o dispersada, este fenómeno permite la adquisición de datos 3D (*e.g. Imágenes de tomografía computarizada -**CT**, imágenes de resonancia magnética -**MRI**, etc.*)

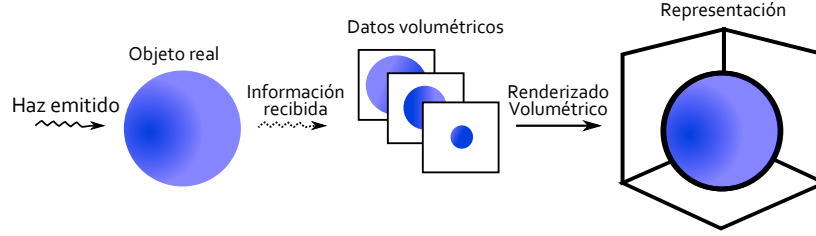


Figura 3.1: Esquema del proceso de la obtención y renderización de los datos volumétricos.

Dado que el modelo completo de interacción de la luz con los objetos no es aplicable para propósitos prácticos debido a su solución, existe variantes simplificadas del modelo que consideran partes del fenómeno de interacción. La variante más común utilizada para el renderizado volumétrico es el de “emisión-absorción”, el cual no toma en consideración la luz dispersada o la iluminación indirecta [16] y se puede expresar por medio de la integral de renderizado volumétrico definida como:

$$I(D) = I_0 e^{\int_{s_0}^D \kappa(t) dt} + \int_{s_0}^D q(s) e^{-\int_s^D \kappa(t) dt} ds \quad (3.1)$$

Donde I_0 es la radiación inicial en una posición de entrada s_0 , la cual es atenuada por el objeto con coeficiente de absorción κ mientras lo atraviesa, $q(s)$ es la emisión que contribuye a la radiación $I(D)$ en un punto de salida D . Este modelo es independiente de la dirección en la que atraviese la luz, y es usado frecuentemente para el análisis de angiografías de resonancia magnética, o la visualización de tomografías de emisión de positrones [32].

3.1. Funciones de transferencia

Una función de transferencia (FT) es una función matemática que asigna un valor de salida correspondiente a cada posible valor de entrada [33]. En óptica, una FT especifica qué tan diferentes son las frecuencias espaciales existentes en el sistema de estudio, en otras palabras, describe cómo los objetos proyectan la luz con la que interactúan [34].

Debido a que los datos volumétricos contienen únicamente valores de intensidad (color y opacidad), cuando se requiere una visualización más detallada o de estructuras específicas, es imposible hacerlo solo con dichos valores de intensidad. Por lo tanto, es esencial plantear un método de mapeo de las intensidades originales a atributos ópticos basados en la información original. Las FT son herramientas de mapeo de datos que permiten acceder a información más

detallada sobre las propiedades de los datos volumétricos. Como resultado, las estructuras de interés pueden ser visualizadas con mayor calidad, asignando en el nuevo espacio de representación generado por la FT valores específicos de color y opacidad [32]. Es importante resaltar que las FT permiten hacer una caracterización sobre el conjunto de datos, asignando valores a cada voxel de los datos. Existen una serie de funciones de transferencia para el renderizado volumétrico que han sido propuestas en el estado del arte para la caracterización de estructuras de interés, muchas de estas de FT son multidimensionales tales como: FT Gradiente, FT Escala, FT Estadística, etc.

3.1.1. Función de transferencia Gradiente

Dado que las primeras FT presentadas en el estudio del renderizado volumétrico fueron unidimensionales y solo asignan valores de opacidad a ciertos voxels en un un histograma de intensidad, la tarea de resaltar información específica es complicada. Kindlmann y Durkin [35] proponen que las características de interés en los datos 3D son los bordes entre objetos o áreas de material relativamente homogéneo, lo que permite asignar valores de opacidad a un rango de voxels que pueden pertenecer al mismo tipo de material, o incluso poder visualizar las superficies que los separan. Este método es especialmente útil cuando el ruido afecta a los datos, haciendo que las fronteras entre los objetos sean confusas como lo muestra la figura 3.2.

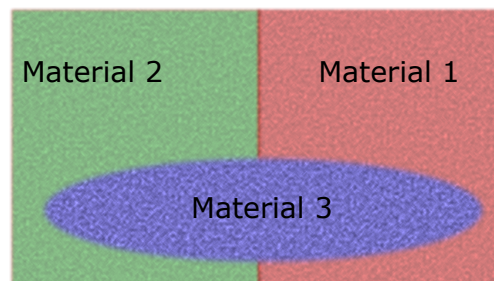


Figura 3.2: Datos corruptos por el ruido.

Puesto que el objetivo es visualizar las fronteras entre los materiales, se formula el modelo de bordes, que asume que las fronteras de los objetos están demarcadas. Sin embargo, es frecuente que las fronteras de estructuras biológicas en los datos volumétricos sean difusas, a causa de propiedades físicas del proceso de captura (ver figura 3.3).

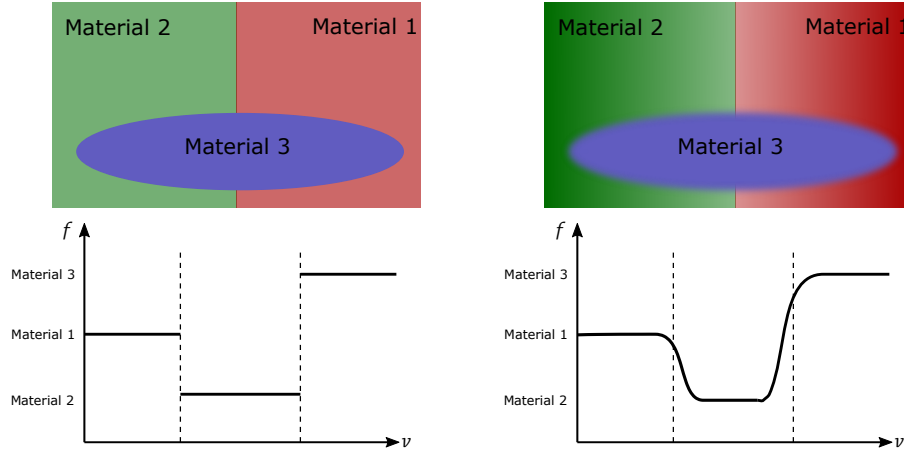


Figura 3.3: Comparación entre los bordes ideales y observados en una imagen volumétrica.

Tomando como base la curva resultante de la integral de una Gaussiana, llamada «función de error» [36], los cambios atenuados hacen que caracterizar los objetos sea más complicado, puesto que se hace más difícil delimitar los objetos. Aplicando las propiedades matemáticas del vector gradiente en una posición cualquiera que siempre apunte perpendicular a una iso-superficie en ese punto cualquiera, se usa el vector gradiente como una manera para encontrar la dirección de concentración del material.

La derivada direccional de un campo escalar f a lo largo de un vector \mathbf{v} , denotada como $D_{\mathbf{v}}f$, es la derivada de f como si se moviese a lo largo de la dirección de \mathbf{v} . Tomando f como una función de una variable, y tomando en cuenta que el eje que se analizará siempre sigue $\widehat{\nabla f} = \mathbf{v}$, los constantes cambios de orientación dependerán de la posición. Esta función se calcula mediante la definición de derivada, es decir, calculando la magnitud del gradiente. Matemáticamente se define como [37]:

$$\begin{aligned}
 D_{\mathbf{v}}f &= \nabla f \cdot \mathbf{v} \\
 D_{\widehat{\nabla f}}f &= \nabla f \cdot \widehat{\nabla f} = \nabla f \cdot \frac{\nabla f}{\|\nabla f\|} = \|\nabla f\| \\
 \|\nabla f\| &= \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 + \left(\frac{\partial f}{\partial z}\right)^2}
 \end{aligned} \tag{3.2}$$

En el caso de la ecuación (3.2), dado los datos 3D, la función f está en función de las direcciones x , y y z . Como f , para este caso, es un conjunto discreto de datos volumétricos, se puede calcular mediante la convolución de un volumen con un Kernel $n \times n$ unidimensional (sobre cada eje espacial). Este proceso se muestra en la figura 3.4, donde cada voxel del Kernel (Coloreados) se alinea sobre una sección del volumen original, y se multiplican uno a uno los elementos correspondientes, posteriormente se suman cada uno de ellos y el valor resultante

es asignado al voxel central (coloreado rojo), el proceso se repite moviendo el Kernel sobre cada uno de los voxels del volumen original. Ya que los lados del Kernel tienen un tamaño de n voxels, para aquellos voxels que se encuentran en el frontera del volumen original, se pueden aplicar técnicas para calcular estos valores, como añadir otra capa con los mismos valores o añadir una capa de ceros, etc.

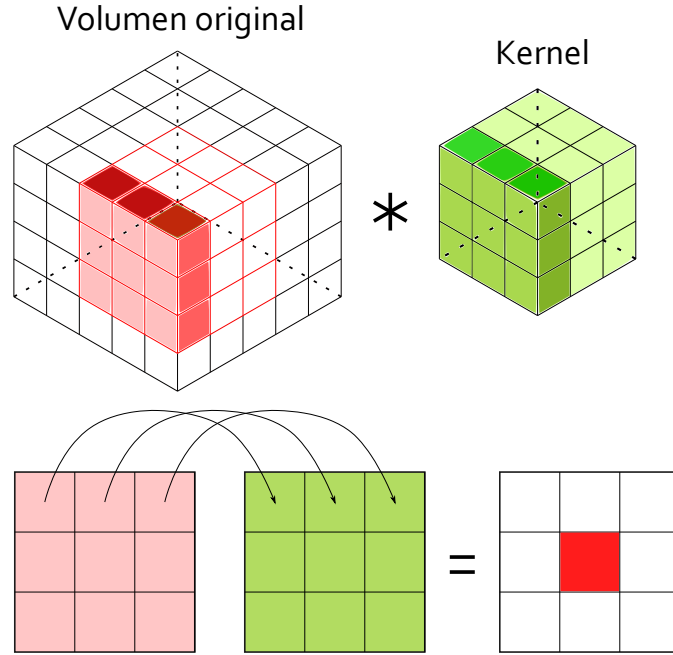


Figura 3.4: Ilustración de la derivación mediante un Kernel usando convolución.

El Kernel utilizado para el cálculo de las derivadas sobre cada dirección, es el operador de tamaño $3 \times 3 \times 3$ cuyos valores se determinaron como un gradiente esférico, que resalta el dato central y da menos importancia aquellos valores alejados del central:

$$\begin{aligned}
 Sx_x &= \begin{Bmatrix} -1 & 0 & 1 \\ -3 & 0 & 3 \\ -1 & 0 & 1 \end{Bmatrix} & Sy_x &= \begin{Bmatrix} -1 & -3 & -1 \\ -3 & -6 & -3 \\ -1 & -3 & -1 \end{Bmatrix} & Sz_x &= \begin{Bmatrix} -1 & -3 & -1 \\ 0 & 0 & 0 \\ 1 & 3 & 1 \end{Bmatrix} \\
 Sx_y &= \begin{Bmatrix} -3 & 0 & 3 \\ -6 & 0 & 6 \\ -3 & 0 & 3 \end{Bmatrix} & Sy_y &= \begin{Bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{Bmatrix} & Sz_y &= \begin{Bmatrix} -3 & -6 & -3 \\ 0 & 0 & 0 \\ 3 & 6 & 3 \end{Bmatrix} \\
 Sx_z &= \begin{Bmatrix} -1 & 0 & 1 \\ -3 & 0 & 3 \\ -1 & 0 & 1 \end{Bmatrix} & Sy_z &= \begin{Bmatrix} 1 & 3 & 1 \\ 3 & 6 & 3 \\ 1 & 3 & 1 \end{Bmatrix} & Sz_z &= \begin{Bmatrix} -1 & -3 & -1 \\ 0 & 0 & 0 \\ 1 & 3 & 1 \end{Bmatrix}
 \end{aligned}$$

La Figura 3.5 muestra gráficamente la aplicación de la FT Gradiente para uno de los volúmenes más utilizados en el estado del arte [35], [13].

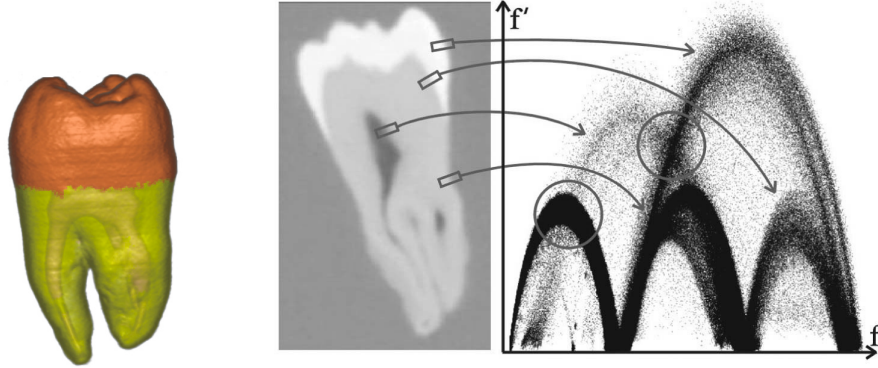


Figura 3.5: Implementación FT gradiente.

3.1.2. Función de transferencia Laplaciana

La FT Laplaciana o de segunda derivada propuesta también por [35], se plantea como respuesta al inconveniente generado en las fronteras de los materiales con la FT de gradiente, el cual hace difícil asociar voxels en las fronteras a sus respectivos materiales. Es decir, existen ciertas secciones de los arcos descritos por la FT Gradiente, o incluso arcos completos que se ven solapados con otros, y esto no necesariamente expresa que ambos bordes pertenecen al mismo objeto o que sean vecinos. Por lo tanto, siguiendo la misma metodología derivativa de la FT Gradiente basada en la información de la segunda derivada, la cual posee información necesaria para distinguir los arcos formados por la primera derivada, en la FT Laplaciana es posible diferenciar entre arcos *positivos* o de curva positiva, y arcos *negativos*, permitiendo diferenciar bordes de diversos objetos.

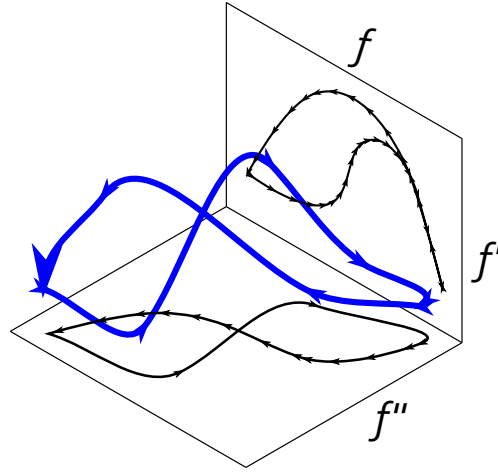


Figura 3.6: Separación de intensidades dada la información direccional de la segunda derivada.

Por tanto, se procede hallando el gradiente del gradiente [32, 35]:

$$D_{\nabla f}^2 f = \frac{1}{\|\nabla f\|} \nabla (\|\nabla f\|) \nabla f \quad (3.3)$$

Este cálculo sin embargo, involucra determinar la matriz Hessiana (H) para cada voxel:

$$\begin{aligned} D_{\nabla f}^2 f &= \frac{1}{\|\nabla f\|^2} \nabla f^\top H f \nabla f \\ &= \frac{1}{\|\nabla f\|^2} \begin{bmatrix} \frac{\partial f}{\partial x} & \frac{\partial f}{\partial y} & \frac{\partial f}{\partial z} \end{bmatrix} \begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial x \partial z} \\ \frac{\partial^2 f}{\partial x \partial y} & \frac{\partial^2 f}{\partial y^2} & \frac{\partial^2 f}{\partial z \partial y} \\ \frac{\partial^2 f}{\partial x \partial z} & \frac{\partial^2 f}{\partial z \partial y} & \frac{\partial^2 f}{\partial z^2} \end{bmatrix} \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \\ \frac{\partial f}{\partial z} \end{bmatrix} \end{aligned} \quad (3.4)$$

aunque la FT Laplaciana solo implica el cálculo de la traza de la matriz Hessiana.

3.1.3. Función de transferencia Valores LH

Las FT Gradiente y FT Laplaciana, extraen la información del vecindario de cada voxel, dando resultado a un espacio donde los bordes pueden ser separados, además de separar estructuras de alto contraste. A pesar de esto, estas FT, basadas en derivadas, presentan un problema en cuanto a caracterizar detalles superficiales que no comparten relación espacial alguna, debido a que la selección de estos voxeles solo se mide en relación a la cantidad de valores por unidad de volumen y no por la intensidad de los voxeles. En otras palabras, varios objetos o estructuras en los datos 3D pueden llegar a verse como un solo punto de concentración. Por esto, en [32] se propone la FT de Valores Bajos(**L**)/Altos (**H**) que tiene como

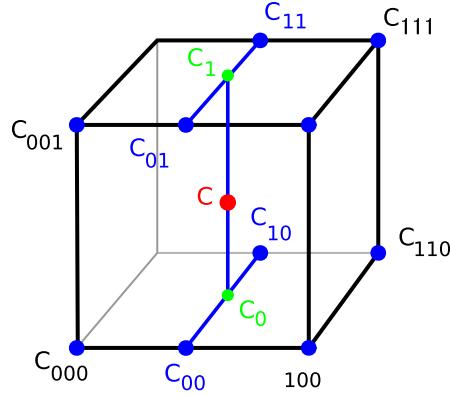


Figura 3.7: Descripción gráfica de la interpolación trilineal.

finalidad mantener las propiedades de las FT Gradiente y FT Laplaciana, donde se pueden ver las transiciones entre materiales y su dirección de concentración, y además, diferenciar las estructuras presentes en el volumen.

Con base en la FT Gradiente y FT Laplaciana, la FT de Valores LH propone utilizar estos volúmenes de representación generados por ambas FT, y usar un método de integración paso a paso, para asignar a aquellos voxels que se encuentran en las fronteras entre materiales, las intensidades de aquellos voxels con mayor intensidad y menor intensidad a los más cercanos a estos que se encuentren en una zona homogénea (*Voxels al interior de un objeto*). Con el fin de hacer el recorrido para la búsqueda de los valores LH para un voxel en la frontera, se utiliza el método de integración de Runge-Kutta de segundo orden con dos pasos [38], definido como:

$$y_{i+1} = y_i + \frac{1}{2}h (\nabla f(y_i) + \nabla f(y_i + h\nabla f(y_i))) \quad (3.5)$$

Donde y es el valor de intensidad, h es el paso de búsqueda y $\nabla f(y)$ el valor de gradiente del voxel en la posición y . Este proceso se realiza para cada uno de los voxels en las fronteras, donde el valor H correspondiente está determinado siguiendo el camino en la dirección del gradiente hacia donde crece la concentración, mientras que el valor L se encuentra siguiendo el camino en la dirección negativa del gradiente. Cabe resaltar que las posiciones a lo largo de los caminos no necesariamente corresponden a un voxel del conjunto de datos, es decir, el valor buscado puede estar entre dos que si existen. Por ende, es necesario utilizar un método de interpolación para hallar este valor. Para este cometido, es propuesto un interpolador trilineal definido como:

$$\begin{aligned} x_d &= (x - x_0) / (x_1 - x_0) \\ y_d &= (y - y_0) / (y_1 - y_0) \\ z_d &= (z - z_0) / (z_1 - z_0) \end{aligned}$$

Donde $(x_0, x_1, y_0, y_1, z_0, z_1)$ son las posiciones conocidas, x_d , y_d y z_d son las diferencias entre

ellas y las de interés (x, y, z) . Conocidas las imágenes o intensidades en los puntos definidos, la intención es realizar una interpolación sobre el eje x :

$$\begin{aligned} C_{00} &= C_{000} (1 - x_d) + C_{100} x_d \\ C_{01} &= C_{001} (1 - x_d) + C_{101} x_d \\ C_{10} &= C_{010} (1 - x_d) + C_{110} x_d \\ C_{11} &= C_{011} (1 - x_d) + C_{111} x_d \end{aligned}$$

Donde $C_{i,j,k}$ es la imagen o intensidad en el punto $(x_i, y_j, z_k) \mid i, j, k \in \{0, 1\}$. Posteriormente se necesita interpolar a lo largo del eje y :

$$\begin{aligned} C_0 &= C_{00} (1 - y_d) + C_{10} y_d \\ C_1 &= C_{01} (1 - y_d) + C_{11} y_d \end{aligned}$$

Y por último, realizar una interpolación lineal sobre el eje z :

$$C(x, y, z) = C_0 (1 - z_d) + C_1 z_d \quad (3.6)$$

Donde $C(x, y, z)$ es el valor de intensidad de la posición de interés.

3.1.4. Función de transferencia Curvatura

El modelo general que sigue la FT Gradiente, está basado en usar la dirección del gradiente como entrada para definir su forma, basado en la interacción de la luz con las superficies [39]. Con el fin de evitar los problemas asociados a la representación generada por el gradiente (i.e. el solapamiento de objetos, la dirección de cambio, entre otros), se ha propuesto una nueva familia de FT basadas en la representación de la segunda derivada, más específicamente en el cambio direccional de la concentración de puntos en los datos volumétricos [40].

Hay dos teorías fundamentales tras la representación por curvatura de los datos volumétricos. La primera es la diferencia geométrica; además de que la curvatura de una superficie puede ser calculada directamente, la representación basada en curvatura constituye un espacio donde las estructuras son geoméricamente ortogonales a las superficies descritas como un invariante. El segundo trasfondo teórico yace en el diseño de filtros y el procesamiento; las diferencias numéricas realizadas en el proceso se acentúan sobre el ruido, y el problema tiende a incrementar su orden de complejidad con el orden la derivada.

La curvatura de una superficie está definida por la relación entre los pequeños cambios posicionales presentes en la curvatura, y los cambios resultantes en la normal a la superficie. En datos volumétricos, las superficies son implícitamente representadas como iso-superficies de reconstrucción continua de datos $f(\mathbf{x})$. Asumiendo que los valores de f incrementan a

medida que nos movemos hacia el interior de los objetos de interés, la superficie normal está definida como $\mathbf{n} = \frac{-\mathbf{g}}{|\mathbf{g}|}$, con $\mathbf{g} = \nabla f$. La información de curvatura está contenida en $\nabla \mathbf{n}^\top \in \mathbb{R}^{3 \times 3}$. Sin embargo, debido a que la curvatura se calcula mediante una convolución sobre un volumen, la complejidad al medir el gradiente en cada punto aumenta [41–43].

Expandiendo la expresión se obtiene:

$$\begin{aligned}
 \nabla \mathbf{n}^\top &= -\nabla \left(\frac{\mathbf{g}^\top}{|\mathbf{g}|} \right) = -\left(\frac{\nabla \mathbf{g}^\top}{|\mathbf{g}|} - \frac{\mathbf{g} \nabla^\top |\mathbf{g}|}{|\mathbf{g}|^2} \right) \\
 &= \frac{-1}{|\mathbf{g}|} \left(\mathbf{H} - \frac{\mathbf{g} \nabla^\top (\mathbf{g}^\top \mathbf{g})^{1/2}}{|\mathbf{g}|} \right) = \frac{-1}{\mathbf{g}} \left(\mathbf{H} - \frac{\mathbf{g} \nabla^\top (\mathbf{g}^\top \mathbf{g})}{2 |\mathbf{g}| (\mathbf{g}^\top \mathbf{g})^{1/2}} \right) \\
 &= \frac{-1}{|\mathbf{g}|} \left(\mathbf{H} - \frac{\mathbf{g} (2 \mathbf{g}^\top \mathbf{H})}{2 |\mathbf{g}|^2} \right) = \frac{-1}{|\mathbf{g}|} \left(\mathbf{I} - \frac{\mathbf{g} \mathbf{g}^\top}{|\mathbf{g}|^2} \right) \mathbf{H} \\
 &= \frac{-1}{|\mathbf{g}|} (\mathbf{I} - \mathbf{n} \mathbf{n}^\top) \mathbf{H}
 \end{aligned} \tag{3.7}$$

Donde \mathbf{I} es la matriz identidad, y \mathbf{H} la matriz Hessiana. Haciendo $\mathbf{P} = \mathbf{I} - \mathbf{n} \mathbf{n}^\top$, se obtiene:

$$\nabla \mathbf{n}^\top = \frac{-\mathbf{P} \mathbf{H}}{|\mathbf{g}|} \tag{3.8}$$

de donde se puede interpretar que \mathbf{H} representa los cambios del gradiente como una función de los cambios infinitesimales de la posición en \mathbb{R}^3 [37]. Los cambios en \mathbf{b} tienen una componente a lo largo de \mathbf{g} (*el gradiente puede cambiar de longitud*), y una componente en el interior del plano tangencial (*el gradiente puede cambiar de dirección*). \mathbf{P} y \mathbf{H} son matrices simétricas, pero $\nabla \mathbf{n}^\top$ no lo es. Sin embargo, si \mathbf{v} está en el plano tangente, entonces $\mathbf{P} \mathbf{v} = \mathbf{v}$ y $\mathbf{v}^\top \mathbf{P} = \mathbf{v}^\top$, así que para \mathbf{u} y \mathbf{v} en el plano tangente:

$$\mathbf{v}^\top \mathbf{P} \mathbf{H} \mathbf{u} = \mathbf{v}^\top \mathbf{H} \mathbf{u} = \mathbf{u}^\top \mathbf{H} \mathbf{v} = \mathbf{v}^\top \mathbf{P} \mathbf{H} \mathbf{v} \tag{3.9}$$

existe una base ortogonal $\{\mathbf{p}_1, \mathbf{p}_2\}$ para el plano tangente en el cual $\nabla \mathbf{n}^\top$ es una matriz diagonal 2×2 , siendo $\nabla \mathbf{n}^\top = \frac{-\mathbf{P} \mathbf{H}}{|\mathbf{g}|}$ la restricción simétrica al plano tangente. Esto puede ser extendido a una base ortogonal en \mathbb{R}^3 , $\{\mathbf{p}_1, \mathbf{p}_2, \mathbf{n}\}$. En esta base la superficie normal derivada es:

$$\nabla \mathbf{n}^\top = \begin{bmatrix} \kappa_1 & 0 & \sigma_1 \\ 0 & \kappa_2 & \sigma_2 \\ 0 & 0 & 0 \end{bmatrix} \tag{3.10}$$

Donde κ_i es la curvatura en sentido positivo ($i = 1$) y en sentido negativo ($i = 2$), σ_i son las desviaciones estándar de cada una. Además la última fila es cero ya que no hay cambios

en la posición que puedan hacer variar la posición de la normal \mathbf{n} . Por último y definiendo $\mathbf{G} = \nabla \mathbf{n}^\top \mathbf{P}$, podemos calcular cada curvatura como:

$$\kappa_{1,2} = \frac{T \pm \sqrt{2F^2 - T^2}}{2} \quad (3.11)$$

Donde $T = \text{tr}(\mathbf{G})$ y $F = \|\mathbf{G}\|_F$.

3.1.5. Función de transferencia Estadística

Con la idea de discriminar estructuras aún cuando tengan valores de intensidad similares, Caban y Rheingans [44] han introducido una técnica basada en el análisis de textura de los datos 3D, combinando estadísticos de primer y segundo orden. El espacio generado a partir de estos estadísticos captura las propiedades estructurales y geométricas del volumen original usando histogramas estadísticos, coocurrencia y matrices de RLE (*Run-Length Encoding*). Inicialmente, el conjunto de datos se divide en subvolumenes solapados que se analizan posteriormente. Estas particiones deben ser lo suficientemente grandes para capturar las propiedades estadísticas de textura y lo suficientemente pequeñas para solamente preservar las características locales alrededor de los voxels.

Un histograma es una medida que mapea valores de intensidad a ciertos bins. Sea $\pi(i)$ con $(i = 1, 2, \dots, n)$, el número de voxels que tienen cierta intensidad i en un sub-volumen dado v_k de tamaño m^3 . La probabilidad de que la intensidad i este dentro del sub-volumen v_k es:

$$h(i) = \frac{\pi(i)}{m^3} \quad (3.12)$$

De donde se puede extraer algunas propiedades estadísticas como la media, la varianza, kurtosis, etc. Los momentos estadísticos de segundo orden miden la verosimilitud de un valor de intensidad i y j en un distancia promedio $\vec{d} = (dx, dy)$ [45]. Los estadísticos de segundo orden se calculan usando matrices de co-ocurrencia. Sea $\delta = (r, \phi)$ un vector en coordenadas polares de un sub-volumen v_k . Se puede calcular la probabilidad conjunta de una pareja de niveles grises que pueden ser separados por δ . Esta probabilidad conjunta se almacena en una matriz $\mathbf{P}(i, j)$ y con ellos posteriormente realizar el cálculo de los momentos estadísticos, que usualmente son la media μ y la desviación estándar σ .

3.1.6. Función de transferencia Escala

La FT Escala ha sido propuesta con la intención de mapear el tamaño local de las características de color y opacidad de cada estructura en los datos, teniendo en cuenta que pueden

tener tamaños diferentes [14,46]. Esta función propuesta por [13] se basa en la obtención local de las *características*, teniendo como eje principal la conservación de su tamaño respecto al volumen en general, donde cada *característica* es un objeto dentro del volumen. En otras palabras, esta FT pretende representar los espacios donde mejor se aprecien las estructuras de acuerdo a su tamaño relativo dentro de la imagen 3D.

Dada una señal N -dimensional continua $f : \mathbb{R}^N \rightarrow \mathbb{R}$, su representación en el espacio de escala continuo es una familia de señales $L : \mathbb{R}^N \times \mathbb{R}_+ \rightarrow \mathbb{R}$ que satisfacen la ecuación de difusión:

$$\partial_t L = \frac{1}{2} \nabla^2 L \quad (3.13)$$

Con condiciones iniciales $L(\mathbf{x}; 0) = f(\mathbf{x})$ el volumen original o en el volumen en escala cero y t la escala. La solución a esta ecuación está dada por:

$$g(\mathbf{x}; t) = \frac{1}{2\pi t} \exp\left(-\frac{\|\mathbf{x}\|^2}{2t}\right) \quad (3.14)$$

La representación del espacio de escala que se busca puede ser obtenida mediante la convolución:

$$L(\mathbf{x}; t) = g(\mathbf{x}, t) * f(\mathbf{x}) \quad (3.15)$$

Sin embargo, este cálculo cuando se aplica a un volumen de (256^3) voxeles, necesitaría 1GB de memoria para almacenarlo, si además queremos encontrar varias representaciones de escala, este volumen tendría n veces su representación por cada escala definida. Por lo tanto, se propone calcular una representación discreta del espacio de escala iterativamente, utilizando integración hacia adelante de Euler:

$$L(\mathbf{x}; t + \Delta t) = L(\mathbf{x}; t) + \Delta t \sum_{ijk \in N} \alpha_{ijk} \Phi(L(\mathbf{x}_{ijk}; t) - L(\mathbf{x}; t)) \quad (3.16)$$

Donde N es un conjunto de índices de los puntos de la vecindad del voxel actual (18 vecinos en este documento), Δt es el paso de cálculo de escala, usualmente 1, y Φ es la función de flujo definida como:

$$\Phi(a) = a \cdot C(a) \quad (3.17)$$

con C la función de conductividad, definida por Perona y Malik, 1990 [47] como una función no lineal basada en la magnitud del gradiente, que puede preservar los bordes:

$$C(\|\nabla L\|) = \frac{1}{1 + \left(\frac{\|\nabla L\|}{\lambda}\right)^2} \quad (3.18)$$

Para λ el parámetro de sensibilidad, el cual, cuando crece C se aproxima a 1 lo que es análogo a una difusión lineal (*interior de un objeto*), y se detiene la difusión cuando se encuentra en los bordes.

Una vez el espacio discreto es calculado, una etapa de selección de escala se hace necesaria, puesto que no todas contendrán información relevante. Esta selección se hace asumiendo máximos locales, puesto que cuando un máximo se halle, esa escala para cierto voxel posee su máxima información, una región de sus vecinos tendrán su máxima información en esa escala [13]. En tal sentido se utiliza el Laplaciano normalizado, y la detección se define como:

$$t\nabla^2 L = t \left(\frac{\partial^2 L}{\partial x^2} + \frac{\partial^2 L}{\partial y^2} + \frac{\partial^2 L}{\partial z^2} \right) = t * \text{tr}(\mathbf{H}) \quad (3.19)$$

Se puede apreciar que cuando el Laplaciano encuentre su valor máximo, esa escala será la representativa, puesto que está normalizado. Por último, dado que se requiere encontrar una representación continua del volumen en el espacio de escala, se hace necesario una retro-proyección que simule la función continua de escala teniendo en cuenta las propiedades del espacio discreto. Debido a que hay características que pueden solaparse, se debe ajustar una escala que exhiba cierto grado de continuidad, y como los datos obtenidos después de la selección de escalas son dispersos, un método rápido para obtener esta escala, es la interpolación de Shepard. Por tanto, el espacio de escala S se define como:

$$S(\mathbf{x}) = \sum_{\mathbf{x}_i \in M} \Theta(\|\mathbf{x} - \mathbf{x}_i\|) t_i \quad (3.20)$$

donde $\Theta(d)$ es la función base, definida como un polinomio de Wendland de cuarto grado:

$$\Theta(d) = ([1 - d/h]_0^1)^4 \left(\frac{4d}{h} + 1 \right) \quad (3.21)$$

donde $[\cdot]_0^1$ es una función de limite entre 0 y 1, y h es un parámetro que controla el resultado de la función base, asumida como $h = kt_i$ con $k = 1$.

Capítulo 4

Análisis de relevancia

Dado que en muchas de las aplicaciones la cantidad de datos es muy abundante (10^6), donde cada observación tiene un número grande de características (> 1000), se conforman bases de datos de basto tamaño que no pueden ser siempre procesadas por los sistemas convencionales de cómputo, haciendo necesario realizar la reducción de dimensión en tales tipos de bases de datos. Este problema ha venido ganando relevancia a través de los últimos años, debido a su importancia en las aplicaciones de aprendizaje de máquinas actuales [48]. Muchos algoritmos y técnicas de reducción de dimensión han planteado encontrar las características de mayor información local para describir o representar un estado, otros lo han hecho buscando esa información en las observaciones, mientras que otros han combinado ambas. En tal sentido en esta sección se describen algunos de los algoritmos más representativos del estado del arte, y comúnmente usados en la etapa de reducción de dimensión para comparar la eficacia de metodologías y algoritmos propuestos.

4.1. Análisis de relevancia estocástico

Un conjunto de características estocástico puede escribirse como una combinación lineal de $D' < D$ funciones bases independientes donde el mínimo del error cuadrático medio se asume como la evaluación medida de un sub-espacio basado en una transformación lineal [49]. El objetivo es llevar el análisis de relevancia basado en variabilidad a un sub-espacio de proyección que maximice la información existente en el espacio de entrada, preservando los datos que más contribuyen a la tarea específica [50, 51]. Dado un conjunto de entrada $\mathbf{X} \in \mathbb{R}^{L \times N}$ donde L es el número de observaciones y N la dimensionalidad de los datos (características), el conjunto de vectores ortogonales es estimado para que los componentes

resultantes de D' puedan aproximar cada característica de entrada en \mathbf{X} de tal manera que la información original se preserve lo máximo posible [52, 53].

Uno de los algoritmos más utilizados en las aplicaciones de reducción de dimensión es el «análisis de componentes principales» o PCA que consiste en tomar una matriz dada $\mathbf{X} \in \mathbb{R}^{N \times P}$ con N muestras o filas y P características o columnas y encontrar un sub-espacio de k dimensiones determinado por la matriz $\mathbf{W} \in \mathbb{R}^{N \times k}$, en el cual la proyección de los datos tienen la máxima varianza posible. Sin embargo, si P tiende a ser grande (> 1000), la exigencia de operaciones crece con el $\min\{NP^2, N^2P\}$ lo que genera una escala de operaciones imposibles para muchas aplicaciones. Por tanto, una alternativa propuesta por O. Shamir [54], consisten en aplicar las ventajas de los métodos estocásticos iterativos de solución con PCA, manteniendo por un lado el número de operaciones logarítmicamente dependiente de la precisión de la tolerancia ϵ , y por otro lado, escalando el número de operaciones a la suma del tamaño N de los datos y de un factor de *eigengap* λ , en lugar del producto $N \times P$.

4.2. Puntuación Laplaciana

La Puntuación Laplaciana (PL) está fundamentada en el Eigenmapeo Laplaciano [55] y la Proyección de Preservación Local [56]. Sea L_r la notación para la puntuación Laplaciana de la r -ésima característica de \mathbf{X} , definida como:

$$L_r = \frac{\tilde{\mathbf{f}}_r^\top \mathbf{L} \tilde{\mathbf{f}}_r}{\tilde{\mathbf{f}}_r^\top \mathbf{D} \tilde{\mathbf{f}}_r} \quad (4.1)$$

Con:

$$\tilde{\mathbf{f}}_r = \mathbf{f}_r - \frac{\mathbf{f}_r^\top \mathbf{D} \mathbf{1}}{\mathbf{1}^\top \mathbf{D} \mathbf{1}} \mathbf{1}$$

Donde \mathbf{f}_r es el vector de muestras de la r -ésima característica:

$$\begin{aligned} \mathbf{f}_r &= [f_{r1}, f_{r2}, \dots, f_{rm}]^\top, \mathbf{D} = \text{diag}(\mathbf{S} \mathbf{1}) \\ \mathbf{1} &= [1, \dots, 1]^\top, \mathbf{L} = \mathbf{D} - \mathbf{S} \end{aligned} \quad (4.2)$$

y \mathbf{S} es la matriz de pesos del modelo de grafos según la estructura local del espacio de datos para cada par de nodos.

Se construye primero un grafo G de vecinos más cercanos con L nodos. El i -ésimo nodo corresponde a \mathbf{x}_i . Posteriormente, se traza una línea entre los nodos i y j si \mathbf{x}_i y \mathbf{x}_j están

cerca. Es decir, \mathbf{x}_i está entre los k vecinos más cercanos de \mathbf{x}_j o en sentido contrario. Cuando se tiene la información de etiquetas, la línea entre los dos nodos puede compartir la misma etiqueta. Si dos nodos i y j están conectados, se pone $S_{ij} = \exp - \frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{t}$, donde t es una constante de ajuste. En caso contrario se pone $S_{ij} = 0$.

4.3. Información mutua

La función de información mutua se define como la distancia de la independencia entre \mathbf{X} y \mathbf{C} medida mediante la divergencia de Kullback-Leiber [24]. Básicamente la información mutua mide la cantidad de información compartida entre un set de datos \mathbf{X} y las etiquetas \mathbf{C} , a través de su nivel de independencia [57]. Esta definición conlleva a la aproximación de Shannon. La información mutua entre los datos y el objetivo o etiquetas, alcanza su valor máximo cuando la variable dependiente u objetivo es perfectamente descrita por el conjunto de características, en este caso la información mutua será igual a la entropía [58–60]. Sin embargo, calcular la información mutua para más de dos variables se vuelve un problema mayor, su grado de dificultad depende del número de variables que hayan implicadas. Debido a esto se han propuesto formas alternativas de calcular el nivel de dependencia. Algunos criterios son la selección de características por información mutua (MIFS) [58] e información mutua conjunta (JMI) [61].

Criterio MIFS

Se define como la relación del nivel de dependencia que existe entre una variable y el vector objetivo (etiquetas) y el nivel total entre esa misma variable y el resto de variables del conjunto. Siendo, el primer término de dependencia la relevancia y el segundo término la redundancia:

$$J(\mathbf{x}_i) = I(\mathbf{x}_i; \mathbf{C}) - \beta \sum_{\mathbf{x}_j \in S} I(\mathbf{x}_i; \mathbf{x}_j) \quad (4.3)$$

Donde \mathbf{x}_i representa la i -ésima característica que es un posible candidato a ser elegido. Las características más relevantes harán que $J(\cdot)$ alcance sus valores más altos y el parámetro β penaliza el término de redundancia, el cual es la información compartida entre la característica evaluada y las demás, a mayor valor tome el término, más redundante será. Fijando el parámetro β en $1/|S|$, donde $|S|$ denota la cardinalidad de S que es el conjunto de índices de las características de una base de datos \mathbf{X} , se puede obtener el criterio de mínima redundancia máxima relevancia (mRMR) [60];

Criterio JMI

De manera similar el criterio de información mutua conjunta se puede interpretar como la dependencia entre una variable y el vector de etiquetas (relevancia), y un término de relación entre las variables. A diferencia del criterio MIFS, el nivel de redundancia entre variables está mediado por la relación del vector de etiquetas. En otras palabras JMI, calcula una relación entre la relevancia y la redundancia entre variables dado la información que compartan simultáneamente con el vector objetivo. Se define así:

$$J(\mathbf{x}_i) = \sum_{\mathbf{x}_j \in S} (I(\mathbf{x}_i; \mathbf{C} | \mathbf{x}_j) + I(\mathbf{x}_j; \mathbf{C})) \quad (4.4)$$

Debido a que el término $I(\mathbf{x}_j; \mathbf{C})$ es una constante en el proceso de selección, la ecuación se puede reescribir de la siguiente manera:

$$J(\mathbf{x}_i) \cong \sum_{\mathbf{x}_j \in S} I(\mathbf{x}_i; \mathbf{C} | \mathbf{x}_j) \quad (4.5)$$

Es importante mencionar que ambos criterios (MIFS y JMI) se pueden generalizar mediante la siguiente ecuación [62]:

$$J(\mathbf{x}_i) = I(\mathbf{x}_i; \mathbf{C}) - \beta \sum_{\mathbf{x}_j \in S} I(\mathbf{x}_j; \mathbf{x}_i) + \lambda \sum_{\mathbf{x}_j \in S} I(\mathbf{x}_j; \mathbf{x}_i | \mathbf{C}) \quad (4.6)$$

Si fijamos ambos parámetros (β y λ) en la ecuación (4.6) a $1/|S|$ podemos obtener el criterio JMI, por otro lado si fijamos λ a 0, podemos obtener el criterio MIFS y finalmente si hacemos $\lambda = 0$ y fijamos $\beta = 1/|S|$ obtenemos la aproximación mRMR. Por lo tanto, la ecuación (4.6) puede ser tomada como forma generalizada para los criterios más populares de la selección de características usando información mutua. Para encontrar una solución óptima usando estos criterios, se debe utilizar un algoritmo que se encargue de obtener el mejor subconjunto de características que maximicen el criterio en cuestión. Sin embargo, como se mencionó anteriormente, algunos de estos criterios se hacen más complejos cuanto mayor sea el número de variables en un conjunto de datos. Por esta razón y con el objetivo de no incrementar la complejidad, este criterio generalizado solo se define para tres variables. Así la información mutua para tres variables, para los criterios más populares, puede ser definida en términos de la entropía ($H(\cdot)$) de la siguiente manera:

$$I(A, B) = H(A) + H(B) - H(A, B) \quad (4.7)$$

$$I(A, B | C) = H(A, C) + H(B, C) - H(A, B, C) - H(C) \quad (4.8)$$

Usando las ecuaciones (4.7) y (4.8) en la ecuación (4.6), se obtiene la ecuación generalizada para MIFS y JMI en términos de entropía:

$$\begin{aligned} J(\mathbf{x}_i) = & H(\mathbf{x}_i) + H(\mathbf{C}) - H(\mathbf{x}_i, \mathbf{C}) - \beta \sum_{\mathbf{x}_j \in S} (H(\mathbf{x}_j) + H(\mathbf{x}_i) - H(\mathbf{x}_i, \mathbf{x}_j)) + \dots \\ & \dots + \lambda \sum_{\mathbf{x}_j \in S} (H(\mathbf{x}_j, \mathbf{C}) + H(\mathbf{x}_i, \mathbf{C}) - H(\mathbf{x}_j, \mathbf{x}_i, \mathbf{C}) - H(\mathbf{C})) \end{aligned} \quad (4.9)$$

No obstante, la estimación de funciones de distribución de probabilidad (PDF) sobre datos continuos (usualmente datos reales) impide, en ocasiones, su aplicación [57]. Sánchez Giraldo *et al.* [63] propone una definición que consiste en encontrar una cantidad que se parezca a la entropía de Renyi en términos del eigenspectro normalizado de la matriz Hermitiana de la proyección de los datos en un espacio de Hilbert (RKHS), obteniendo la entropía directamente de los datos sin estimar la PDF.

4.3.1. Aproximación por entropía de Renyi

Sea $\kappa : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ un Kernel definido positivo evaluado que también es infinitamente divisible. Dada una matriz de entrada $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ y la matriz de Gramm \mathbf{K} obtenida de evaluar un Kernel definido positivo κ sobre todas las parejas posibles de muestras, generando como resultado para las i y j -ésimas muestras \mathbf{x}_i y \mathbf{x}_j , en el elemento $K_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j)$ de la matriz de Gramm \mathbf{K} . De esta manera se puede encontrar una matriz análoga basada en la entropía de orden α para una matriz definida positiva normalizada (DPN) $\mathbf{A} \in \mathbb{R}^{N \times N}$ tal que $\text{tr}(\mathbf{A}) = 1$, puede darse mediante el siguiente funcional:

$$S_\alpha(\mathbf{A}) = \frac{1}{1 - \alpha} \log \text{tr}(\mathbf{A}^\alpha) \quad (4.10)$$

Aquí $S_\alpha(\mathbf{A})$ se aproxima a la entropía de Renyi de \mathbf{A} , tal que $S_\alpha(\mathbf{A}) \approx H_\alpha(\mathbf{A})$, cada elemento de \mathbf{A} se puede definir como: $A_{ij} = \frac{K_{ij}}{N\sqrt{K_{ii}K_{jj}}}$. Análogamente, esta definición permite calcular la entropía de Renyi para más matrices de Gramm:

$$S_\alpha \left(\frac{\mathbf{A} \circ \mathbf{B}}{\text{tr}(\mathbf{A} \circ \mathbf{B})} \right) = \frac{1}{1 - \alpha} \log \text{tr} \left(\left(\frac{\mathbf{A} \circ \mathbf{B}}{\text{tr}(\mathbf{A} \circ \mathbf{B})} \right)^\alpha \right) \quad (4.11)$$

Donde \circ es el producto Hadamard, y $\text{tr}(\cdot)$ es el operador traza.

4.4. Estrategias de búsqueda secuencial

Las estrategias de búsqueda secuencial construyen un conjunto de características, el cual alcanza el valor más alto o bajo, según una función de costo que evalúa el conjunto. Esta clase de estrategias añaden o eliminan características de manera iterativa, y evalúa las combinaciones formadas en cada iteración usando una función de costo. Los algoritmos de búsqueda heurística más comunes son la búsqueda hacia adelante (*FS* “Forward Selection”) y la eliminación hacia atrás (*BE* “Backward Elimination”). Estas estrategias de búsqueda intentan encontrar un subconjunto de características en el espacio de subconjuntos de características de 2^P miembros cuadrados (término \times término) que optimizan la función de costo J . El siguiente algoritmo describe como funcionan estas dos estrategias de búsqueda secuencial. Dada una función de costo J , P características χ_j y un vector de etiquetas \mathbf{c} , seleccionar un subconjunto de $L \leq P$ características que maximicen la función de medidas.

Entrada: Sea $\mathbf{X} \in \mathbb{R}^{N \times P}$, $\mathbf{X} = \{\chi_1, \chi_2, \dots, \chi_P\}$ un conjunto de datos con N muestras y P características.

Salida: Un conjunto S_{opt} de índices que maximizan una función de costo $J(\cdot)$.

Inicialización: $S_0 = \emptyset$ como el conjunto vacío para FS y $S_0 = S_P$ como el conjunto completo para BE, $i = 1$ como el contador de iteraciones, $J(S_0)$ a $-\infty$, se crea un S_{temp} y se inicializa $S_{opt} = \emptyset$ para *FS* y $S_{opt} = S_P$ para *BE*.

Mientras: $J(S_i) > J(S_{i-1})$

 Actualizar $S_{temp} = S_{opt}$

Para: $j = 1 : P - |S_{opt}|$, $j \notin S_{opt}$ características; **Hacer:**

 Añadir la característica j al conjunto temporal $S_i = \{S_{temp} + j\}$.

 Remover la característica j al conjunto temporal $S_i = \{S_{temp} - j\}$.

 Evaluar el conjunto temporal a través de la función de costo $J(S_i)$ y guardar su valor.

 Elegir la característica j que maximiza la función de costo J en los valores guardados para la iteración actual i , j_i .

Si: $J(S_i) > J(S_{i-1})$

 Actualizar $S_{opt} = \{S_{opt}, j_i\}$.

De lo contrario:

 Termine el ciclo.

De manera similar se emplea la búsqueda flotante. Teniendo como mayor diferencia que se aplica inicialmente una búsqueda secuencial hacia adelante hasta encontrar un punto convergencia y con el set encontrado en ese punto, se ejecuta una eliminación hacia atrás. Permitiendo eliminar posibles características que no hacen un aporte eficaz a la optimización de la función de costo.

4.4.1. Algoritmo COBRA

La optimización del problema de selección de características usando los criterios basados en información mutua, se entiende como la tarea de encontrar una combinación óptima de características que maximice los criterios, este planteamiento se entiende como el problema de máximo corte en la teoría de grafos que se puede simplificar al utilizar programación semidefinida positiva. El siguiente algoritmo describe el funcionamiento de la estrategia de búsqueda propuesta por Naghibi *et al.* [57] llamado COBRA. La estrategia está resumida en tres pasos fundamentales.

1. **Problema semidefinido:** Solucionar el problema semidefinido \mathbf{Y}_{SDP} . Repetir el siguiente proceso las veces que se requieran para obtener como salida la mejor solución.
2. **Aproximación aleatoria:** Usando la distribución normal multivariada con media cero y matriz de covarianza $\Sigma_{sdp} = \mathbf{Y}_{SDP}$ para muestrear \mathbf{u} de la distribución $\mathcal{N}(0, \Sigma_{sdp})$ y construir $\tilde{\mathbf{x}} = \text{sign}(\mathbf{u})$. Seleccionar $\mathbf{X} = \{\chi_i | \tilde{x}_i = \tilde{x}_0\}$.
3. **Ajuste de cardinalidad:** Utilizando las estrategias de búsqueda FS o BE re-calcular el tamaño de la cardinalidad de la matriz \mathbf{X} a L .

El paso de ajuste de la cardinalidad es un proceso estándar para generar soluciones binarias de una solución de valor real de un problema semidefinido, y es ampliamente usado para el diseño y análisis de algoritmos de aproximación. En resumen este paso se trata de construir una solución factible que satisfaga la restricción de cardinalidad. Generalmente, este paso puede ser omitido si en el problema de selección la restricción de cardinalidad no requiere la solución exacta.

Parte III

Marco Experimental

Capítulo 5

Esquema general de trabajo

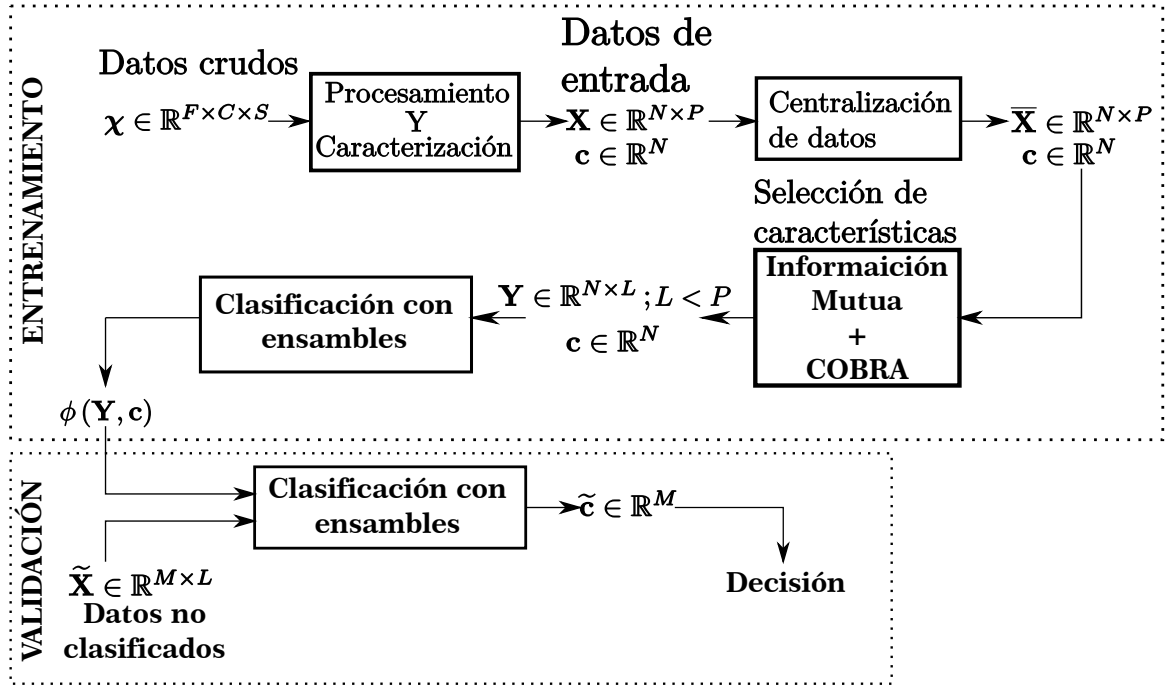


Figura 5.1: Esquema general de la propuesta experimental de trabajo.

Este capítulo tiene como propósito describir cómo se configuran los experimentos de este trabajo (ver Figura 5.1), iniciando con el procesamiento de los datos crudos, caracterización de imágenes MRI en el caso de la aplicación de detección de DCF, hasta la etapa de selección de características, clasificación y posterior validación.

Con el objetivo de evaluar el desempeño de los métodos propuestos de selección de características, el experimento se lleva a cabo sobre datos sintéticos. Se evalúa la eficiencia de la selección mediante la metodología presentada, comparando con las técnicas de selección de características convencionales basadas en variabilidad y búsqueda heurísticas, que son referente en el análisis de relevancia. Posteriormente y haciendo uso de datos sintéticos o controlados, se prueba cada etapa para evaluar su eficiencia en labores de clasificación.

Es importante recalcar que todos los algoritmos utilizados en esta metodología fueron implementados en las herramientas de programación Matlab[®] y Python 3.6, en un equipo con sistema operativo Windows 7, con un CPU Intel[®] Core[™] i7-4900MQ y con 8 Gb de memoria RAM.

5.1. Descripción de bases de datos

En este trabajo se utilizan dos tipos de bases de datos; el primer tipo de base de datos son imágenes de MRI para la detección de DCF. Los resultados obtenidos están directamente relacionados con el objetivo principal de este trabajo. El segundo tipo de bases de datos, se emplearon con la finalidad de evaluar la competitividad de los métodos de análisis de relevancia utilizados en el estado del arte y lo propuesto en esta investigación. Al segundo tipo de base de datos no se aplica la etapa de caracterización por FT. Sin embargo, se utilizan tanto en el análisis de relevancia como en la etapa de clasificación datos.

La Tabla 5.1 describe la información de los pacientes de DCF, en ella se presenta las dimensiones de las imágenes MRI, el número total de voxels y el tipo de DCF.

Paciente ID	Dimensiones $f \times c \times s$	# voxels	Tipo de DCF
1	$168 \times 512 \times 512$	44040192	Temporal derecha
2	$160 \times 512 \times 512$	41943040	Frontal
3	$120 \times 512 \times 512$	31457280	Ínsula izquierda

Tabla 5.1: Base de datos pacientes con DCF.

La Tabla 5.2 describe el segundo tipo de bases de datos, reportadas en el estado del arte [13, 40, 57, 60], empleadas especialmente en la etapa de selección de características. Sin embargo, también se usan para evaluar los resultados de clasificación. La Tabla 5.2 muestra para cada una de las bases de datos el número de muestras N , de características P , el número de clases C , si existen datos faltantes, si existe desbalance, y el orden de desbalance entre clase mayoritaria y minoritaria.

Nombre	# Muestras	# Características	# Clases	¿Datos faltantes?	¿Desbalance?	% C^+ : % C^-
Cardiac Arrhythmia	452	279	13	Sí	Sí	54,2 : 0,04
Australian Credit Approval	690	14	2	Sí	Sí	55,5 : 44,5
Breast Tissue	106	9	6	No	Sí	20,7 : 13,2
Dermatology	366	34	6	Sí	Sí	24,8 : 5,5
German Credit	1000	24	2	No	No	
Glass Identification	214	10	6	No	Sí	32,7 : 4,2
Iris Plants	150	4	3	No	No	
LIBRAS Movement	360	90	15	No	No	
Indian Liver Patient	583	10	2	No	Sí	71,3 : 28,4
Lung Cancer	32	56	3	Sí	Sí	40,6 : 28,1
Parkinson Disease	197	23	2	No	No	
Pima Indians Diabetes	768	8	2	No	Sí	65,1 : 34,9
Sonar, Mines vs. Rocks	208	60	2	No	Sí	54 : 46
Wine Recognition	178	12	3	No	Sí	39,9 : 26,9
Protein Localization Sites (Yeast)	1484	8	10	No	Sí	31,2 : 0,33

Tabla 5.2: Bases de datos utilizadas en el estado del arte.

5.2. Procesamiento y caracterización

Muchas aplicaciones, en distintas áreas de investigación, a menudo involucran datos que inicialmente pueden ser complicados de analizar y utilizar debido al ruido, a datos atípicos, espurios, faltantes e incluso a la misma naturaleza de estos.

Como se menciona en el Capítulo 3, el renderizado volumétrico juega una labor importante en la tarea de caracterización de imágenes de resonancia magnética, las técnicas referentes para de este tipo de imágenes están basadas en el tratamiento de superficies en regiones de interés mediante la comparación directa con un atlas u otras imágenes que contengan la información de referencia, esto se logra a través del uso de software especializado para el registro y procesamiento de la imagen, el cual retorna un conjunto de datos representados en vértices calculados sobre la superficie en la zona de interés. En la metodología propuesta se plantea caracterizar y procesar las imágenes mediante técnicas volumétricas en lugar de las técnicas superficiales convencionales.

Con el fin de evaluar la representación de las imágenes de resonancia magnética, se calcula la FT Gradiente (ver Sección 3.1.1) la cual para cada voxel genera dos valores uno de intensidad y otro de gradiente; la FT Laplaciana (ver Sección 3.1.2) que genera para cada voxel un valor de concentración; la FT Curvatura (ver Sección 3.1.4) que genera para cada voxel dos valores de curvatura, uno en sentido positivo y otro en sentido negativo. Estas tres FT pertenecen al mismo método de caracterización basado en derivación. Se calcula también la FT LH descrita en la Sección 3.1.3, que genera dos valores a cada voxel asociados con sus alrededores; la FT Estadística descrita en la Sección 3.1.5 se basa en asignar un valor de media y desviación estándar a cada voxel según una vecindad de puntos; por último se encuentra la FT Escala (ver Sección 3.1.6) en la que a cada voxel se asigna un valor de escala

representativo según su tamaño local. El conjunto formado por los 10 valores que generan las FT describen en tres dimensiones la MRI, permitiendo conservar la ubicación espacial de la imagen. Adicionalmente, durante el cálculo de estos valores, se ejecuta de manera simultanea un filtrado sobre la MRI para evitar los datos que no pertenecen al volumen de interés, permitiendo que datos de aire y hueso puedan ser excluidos del cálculo, agilizando la caracterización. Cada una de estas técnicas fueron aplicadas a los datos crudos descritos en la Tabla 5.1.

Con el conjunto de datos generado mediante las FT para cada paciente descrito en la Tabla 5.1, se realiza centralización de datos ($\mu = 0$) y estandarización de los mismos ($\sigma = 1$), con el propósito de operar sobre la matriz de correlaciones de los datos, evitando el sesgo por escala

5.3. Selección de características

Sea $\mathbf{X} \in \mathbb{R}^{N \times P}$, $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N\}$ una base de datos con N observaciones y P características, y sea $\mathbf{x}_i \in \mathbb{R}^P, i = 1, 2, \dots, N$ un vector de caracterización P -dimensional, del mismo modo sea $\mathbf{x}_j \in \mathbb{R}^N, j = 1, 2, \dots, P$ un vector de muestras N -dimensional el cual representa la j -ésima característica. Sea $\mathbf{c} \in \mathbb{R}^N$ un vector de etiquetas multi-clase. Los métodos de selección intentan encontrar un subconjunto $\hat{\mathbf{X}} \subseteq \mathbf{X} | \hat{\mathbf{X}} \in \mathbb{R}^{N \times L}$ donde $L \leq P$, este nuevo subconjunto $\hat{\mathbf{X}}$ tiene menos o igual número de características que el conjunto original.

Según lo descrito en el Capítulo 4, la selección de características se realiza a través de cada método mencionado. Para los métodos de selección por búsqueda heurística se requiere de una función de costo para determinar el mejor conjunto. En el caso específico de la FS, BE y FSBE se utilizó la media geométrica debido a que la mayoría de bases de datos presentan desbalance.

Con respecto a la selección usando información mutua, los criterios utilizados para evaluar la explicabilidad de los datos sobre las etiquetas de entrada son la *información mutua conjunta* (JMI) y el criterio de *mínima redundancia máxima relevancia* (mRMR), ambas descritas en la Sección 4.3.

5.3.1. Selección mediante información mutua

Sea S un conjunto de índices i de k características $S = \{1, 2, \dots, k\}$. Dada una base de datos $\mathbf{X} \in \mathbb{R}^{N \times P}$, y un vector de clases \mathbf{C} , la información mutua $I(\cdot, \cdot)$ se define de la siguiente manera:

$$\begin{aligned} I(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k; \mathbf{C}) &= I(\mathbf{X}; \mathbf{C}) \\ &= \int P(\mathbf{X}, \mathbf{C}) \log \frac{P(\mathbf{X}, \mathbf{C})}{P(\mathbf{X})P(\mathbf{C})} d\mathbf{X} d\mathbf{C} \end{aligned} \quad (5.1)$$

Donde $I(\cdot)$ es operador de información mutua, \mathbf{X} es la matriz de datos de entrada, \mathbf{C} es el vector objetivos o de etiquetas y $P(\cdot)$ es la probabilidad.

$$J(\zeta_k) = I(\zeta_k; \mathbf{c}) - \beta \sum_{\zeta_j \in \mathbf{S}} I(\zeta_j; \zeta_k) + \lambda \sum_{\zeta_j \in \mathbf{S}} I(\zeta_j; \zeta_k | \mathbf{c})$$

Siendo $J(\cdot)$ el criterio general de MIFS y JMI, ζ_b la b -ésima característica de la matriz de entrada, y β y λ los parámetros de ponderación de la redundancia.

Los parámetros β y λ controlan la relación entre la relevancia y la redundancia. Es decir, estos parámetros determinan la eficiencia de un conjunto de características con respecto a las etiquetas, manteniendo la información más relevante y menos redundante. El proceso de sintonización de los parámetros se describe de la siguiente manera:

1. Se analiza el comportamiento mediante un experimento sintético de los parámetros según los datos y las etiquetas de entrada. Encontrando un patrón de relación.
2. Comprobado que es posible establecer un patrón definido entre los datos y las etiquetas con los parámetros de la ecuación, se llevo a cabo un experimento controlado con datos reales para determinar qué tipo de modelos se pueden implementar.
3. La ecuación (5.2) se establece para el cálculo del parámetro

$$\beta(\mathbf{X}, \mathbf{c}) = \frac{\sum_{i=1}^P \sum_{j=i+1}^{P-1} H(\zeta_i, \zeta_j)}{\sum_{i=1}^P H(\zeta_i), \mathbf{c} + P \cdot H(\mathbf{c})} \quad (5.2)$$

con $H(\cdot)$ el operador entropía.

El objetivo principal de la ecuación (5.2) es calcular de manera automática el parámetro y que sea dependiente de los datos de entrenamiento, permitiendo además, un relación eficiente entre la relevancia del conjunto de variables y el objetivo.

Una vez definida la ecuación (5.2), el siguiente inconveniente en la selección por información mutua es la evaluación de los diferentes conjuntos de características. Por ello, es necesario el uso de algoritmos de optimización que se basen en una medida de rendimiento sobre el conjunto \mathbf{S} . Las medidas utilizadas son los criterios antes mencionados (JMI y mRMR) y los algoritmos de optimización son FS, BE y COBRA, este último descrito en la Sección 4.4.1.

5.3.2. Selección por métodos de puntuación

Los métodos de selección por puntuación descritos en el Capítulo 4, se basan en asignar un escalafón por puntuación de las variables, usualmente basado en su variabilidad. Las técnicas PCA, PL y Relief, se llevan a cabo sobre los datos centralizados y estandarizados para evitar que características con varianzas y escalas más grandes tengan mayor peso de ponderación en la selección. En el caso de las técnicas de Relief y PL, debido a que las bases de datos de la Tabla 5.2 tienen muchas muestras (> 10000), se debió implementar una aproximación truncada de los métodos de selección.

5.4. Clasificación

La validación utilizada para probar la metodología descrita de clasificación fue una partición hold-out de 70 % de las muestras para entrenamiento y el 30 % restante para prueba. Los clasificadores utilizados para evaluar el desempeño de la selección de características descrita en la Sección 4.3, fueron: clasificador lineal (LDA), árbol de decisión (DT), k-nn de 5 vecinos y máquina de vectores de soporte (SVM), es importante aclarar que ninguno de los parámetros presentes en los clasificadores fue manipulado o sintonizado, es decir, se utilizó la configuración por defecto existente en la copia de Matlab 2017 utilizada en esta investigación. Estos clasificadores también son utilizados como *WeakLearners* en el algoritmo RUSBoost para construir el ensamble en los casos de clasificación desbalanceada. Cada uno de estos clasificadores se implementan en las condiciones más simples y estandarizadas posibles, es decir, no se hace ningún tipo de sintonización de parámetros a ninguno de ellos. Se resalta que estos clasificadores fueron empleados debido a que se requería probar la potencia de reducción de los métodos de análisis de relevancia.

5.4.1. Clasificación de datos desbalanceados

En términos de clasificación, cualquier conjunto de datos que exhiba una distribución desigual entre sus clases puede considerarse desbalanceado. Este fenómeno se produce cuando hay muchas más muestras en una clase que en la otra clase en un conjunto de datos de entrenamiento. En un conjunto de datos desbalanceado, la clase mayoritaria tiene un gran porcentaje de todas las muestras, mientras que las instancias de la clase minoritaria sólo ocupan una pequeña parte de las observaciones.

Por lo general, sin consideración del problema del desbalance de clases, un algoritmo de clasificación tenderá a predecir que las muestras desconocidas pertenecen a la clase mayoritaria e ignorar la clase minoritaria. Sin embargo, en muchas de las aplicaciones, la clase minoritaria

es de vital importancia. Un ejemplo clásico es la clasificación de pacientes con cáncer o sanos a partir de imágenes de mamografías. Basados en la experiencia, el número de pacientes sanos supera notablemente el número de pacientes con cáncer. En consecuencia, en muchos algoritmos estándar de aprendizaje, se encuentran clasificadores que proveen un grado severamente desbalanceado de acierto, con efectividades de casi el 100 % para la clase mayoritaria pero efectividades entre 0 % y 10 % en la clase minoritaria, aún cuando el hecho de saber si el paciente tiene cáncer es fundamental [64, 65].

Se debe resaltar que en la identificación de estructuras cerebrales (*e.g.* DCF), el desbalance es un problema presente debido a la notoria diferencia entre el tamaño local de una estructura específica y el resto de tejido y hueso presentes en las MRI.

Dada la gran importancia del problema del desbalance de clases, distintos algoritmos y métodos han sido propuestos en la última década [66–68]. En tal sentido, se requiere un clasificador que proporcione alta precisión para la clase minoritaria sin poner en peligro la precisión de la clase mayoritaria. Para resolver esto se han propuesto tres estrategias básicas:

1. **Métodos de muestreo:** los cuales son técnicas de preproceso que intentan equilibrar las distribuciones al considerar las proporciones representativas de los ejemplos de clase en la distribución.
2. **Métodos de aprendizaje costo-sensitivos:** los cuales consideran los costos asociados con la clasificación errónea de las muestras [64]
3. **Métodos de ensamble:** los cuales consisten en la combinación de dos o más clasificadores.

Uno de los algoritmos que ha demostrado resultados eficientes es el algoritmo propuesto por Seiffer *et al.* [69], llamado *RUSBoost* y se describe como:

Algoritmo RUSBoost

Dado: Conjunto $S = \{(x_1, y_1), \dots, (x_n, y_n)\} \in \mathbf{X}$ de muestras con clase minoritaria $y^r \in \mathbf{Y}$, $|\mathbf{Y}| = 2$
 Clasificador Base, *WeakLearn*

Número de interacciones, T

Porcentaje deseado del total de instancias a ser representadas por la clase minoritaria, $M = \frac{1}{m}$

1. Inicializar $D_1(i) = \frac{1}{m}$ para todo i

2. Hacer para $t = 1, 2, \dots, T$

a. Crear un conjunto temporal S'_t con distribución D'_t usando submuestreo aleatorio

b. Entrenar el *WeakLearn* con las muestras S'_t y sus pesos de ponderación D'_t

c. Dar una hipótesis $h_t : \mathbf{X} \times \mathbf{Y} \rightarrow [0, 1]$

d. Calcular la pseudo-perdida (para S y D_t):

$$\epsilon_t = \sum_{(i,y):y_i \neq y} D_t(i) (1 - h_t(x_i, y_i) + h_t(x_i, y))$$

e. Recalcular el peso de actualización:

$$\alpha_t = \frac{\epsilon_t}{1 - \epsilon_t}$$

f. Actualizar D_t :

$$D_{t+1}(i) = D_t(i) \alpha_t^{(1+h_t(x_i, y_i) - h_t(x_i, y: y \neq y_i))}$$

g. Normalizar D_{t+1} : Con $Z_t = \sum_i D_{t+1}(i)$

$$D_{t+1}(i) = \frac{D_{t+1}(i)}{Z_t}$$

3. Dar una hipótesis final:

$$H(x) = \arg \max_{y \in \mathbf{Y}} \sum_{t=1}^T h_t(x, y) \log(\alpha_t^{-1})$$

Este algoritmo fue presentado como una variación del algoritmo SMOTEBoost [70]. El algoritmo RUSBoost introduce la técnica de sub-muestreo aleatorio “*Random Under-Sampling*” o RUS que remueve muestras aleatoriamente de la clase mayoritaria hasta que se tenga un porcentaje de representación deseado y luego utilizar la conocida técnica de AdaBoost para entrenar un ensamble de clasificadores y finalmente dar un modelo que tenga en cuenta el orden de desbalance en los datos.

Con el objetivo de evaluar la efectividad de la clasificación, la convención utilizada en este trabajo, considera la clase minoritaria como la negativa y la clase mayoritaria como la positiva, teniendo en cuenta esta aclaración se presentan los siguientes indicadores de desempeño de clasificación:

$$\text{Acc} = \frac{VP + VN}{T}; \quad \text{Error rate} = 1 - \text{Acc} \quad (5.3)$$

Donde *Acc* es la precisión de clasificación, es decir, que tan correctamente el modelo puede

		Valor real		total
		P	N	
Predicción	P'	Verdaderos Positivos (VP)	Falsos Positivos (FP)	N'_p
	N'	Falsos Negativos (FN)	Verdaderos Negativos (VN)	N'_n
total		N_p	N_n	

Tabla 5.3: Matriz de confusión

categorizar una muestra nueva, y VP , VN , P y N son los verdaderos positivos, verdaderos negativos y el total de casos respectivamente. La Tabla 5.3 muestra la matriz de confusión en un caso de dos clases.

Debido a que la clasificación desbalanceada presenta sesgo hacia la clase mayoritaria, se hace necesario utilizar una medida que refleje la representación de las clases sin importar el desbalance. Por tanto, se adopta la media geométrica (G-mean), definida en la ecuación (5.4) para evaluar el desempeño:

$$\text{G-mean} = \sqrt{\frac{VP}{VP + FN} \times \frac{VN}{VN + FP}} \quad (5.4)$$

Capítulo 6

Resultados

En esta sección se presentan los resultados obtenidos por la metodología de trabajo mostrada en el Capítulo 5. Se ilustran los experimentos llevados a cabo en cada etapa utilizando los métodos expuestos. Debe resaltarse que los resultados de la caracterización sólo fueron aplicados a los datos de MRI descritos en la Tabla 5.1. Por otra parte, las técnicas de selección y clasificación son evaluadas tanto en la base de datos MRI como en las demás descritas en la tabla 5.2.

6.1. Procesamiento y caracterización

Empleando las técnicas para el renderizado volumétrico descritas en la Sección 5.2 sobre las MRI de la Tabla 5.1, las siguientes imágenes muestran los resultados de las FT sobre un corte de uno de los pacientes. En las Figura 6.2 se pueden evidenciar los resultados de caracterización de cada FT aplicada sobre la MRI original del paciente 1, donde se resaltan sus particularidades.

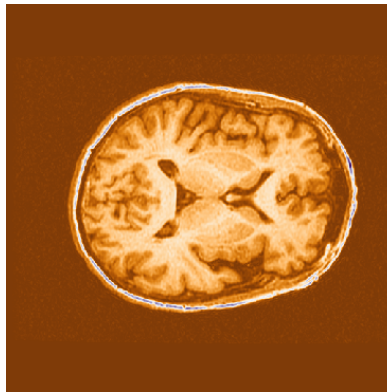


Figura 6.1: Corte original

Con el objetivo de ilustrar los resultados de caracterización de las FT, la Figura 6.1, muestra un corte axial original de la MRI del paciente 1.

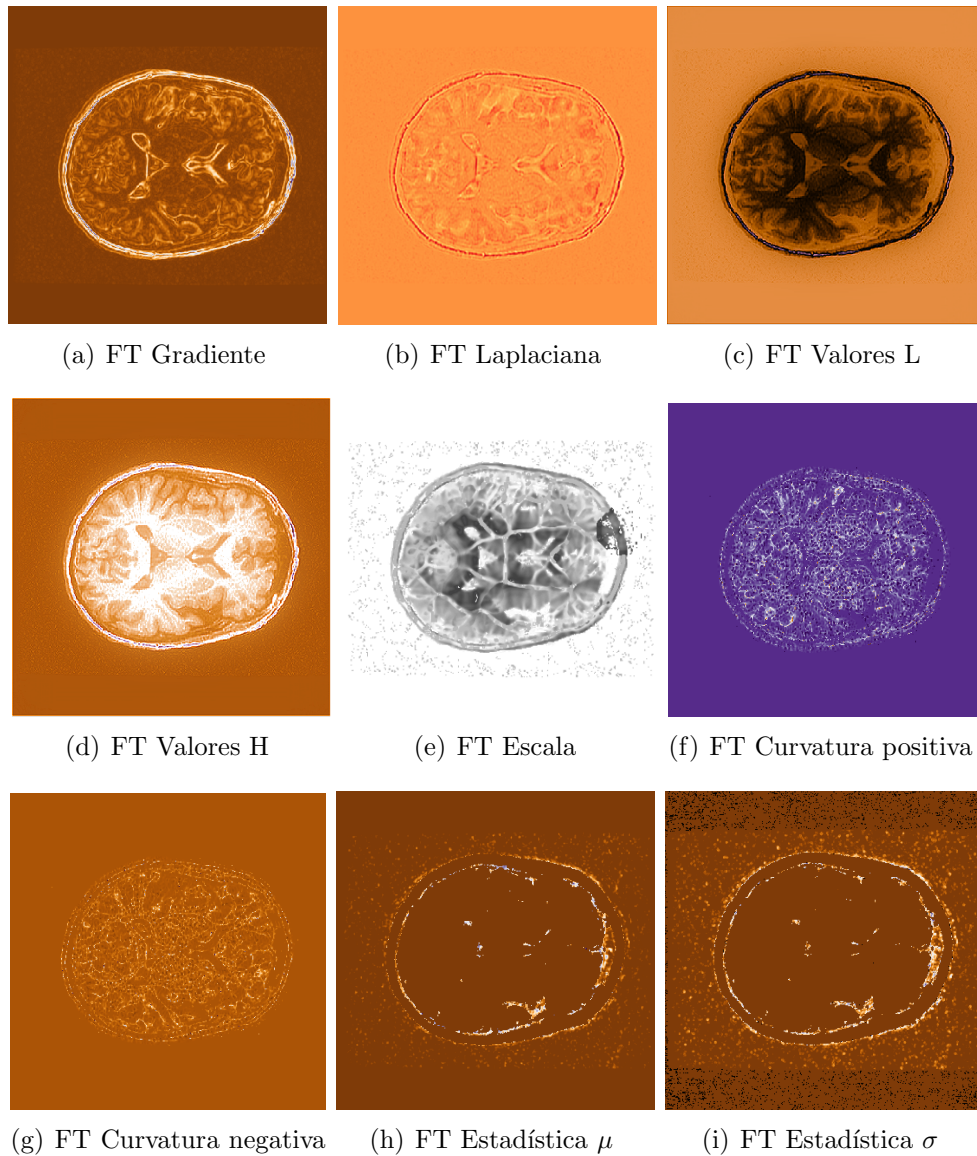


Figura 6.2: Resultados de la caracterización por FT de un corte de MRI.

FT Gradiente: La imagen (6.2(a)) muestra aplicación de ecuación (3.2) como lo menciona la Sección (5.2), en la que se puede detallar el resultado de la operación sobre la imagen, obteniendo como resultado que las zonas de alto contraste pertenecen a los bordes o zonas donde convergen intensidades, y corresponden a materiales diferentes. Esta FT Gradiente es importante en el análisis de estructuras, debido a que, se pueden resaltar cambios de intensidades como por ejemplo, apreciar un detalle claro entre las fronteras de hueso, materia gris y blanca.

FT Laplaciana: La Figura 6.2(b) muestra aplicación de ecuación (3.3) como lo menciona la Sección 5.2, en la que observa el resultado del procesamiento de la MRI, en donde se capta que las fronteras son resaltadas. También pueden determinarse regiones con comportamientos o contrastes similares, permitiendo hallar las zonas de concentración de alta densidad y de baja densidad.

FT Valores LH: Las Figuras 6.2(c) y 6.2(d) muestran la aplicación de ecuación (3.5), representando en la Figura 6.2(c) los valores de cada vóxel asociados a sus vecinos de baja intensidad, y análogamente a en la Figura 6.2(d) aquellos de alta intensidad. En ambas figuras se puede resaltar zonas completas que se atribuyen a vecindades de baja o alta densidad, asignando a regiones con fronteras difusas, una delimitación clara.

FT Curvatura: La Figura 6.2(f) muestra los resultados de aplicación de FT Curvatura en sentido positivo y la Figura 6.2(g) exhibe los resultados para el sentido negativo. En las dos figuras se pueden ver zonas de alto contraste que realzan los cambios en la curvatura asociados a estructuras que tienen cambios abruptos. La FT Curvatura para la aplicación actual presenta los mejores resultados en términos de información aportada para la detección de DCF.

FT Estadística: En cuanto a los resultados de la FT estadística (Figuras 6.2(h) y 6.2(i)), la representación de los momentos de la media μ y desviación σ no exhiben demasiada información, debido a la naturalidad de la FT en sí, que asigna a regiones que comparten propiedades valores cercanos a cero, y exaltando aquellas que no. Aunque ambas representaciones se ven similares es posible notar que existen diferencias en cuanto los valores asignados en las regiones resaltadas.

FT Escala: La Figura 6.2(e) presenta información relacionada al tamaño de los objetos, asignando a los escalas con base en su tamaño local. Por lo tanto, el color blanco se asocia a aquellos objetos en la imagen que tienen el tamaño más grande posible que se puede encontrar. Mientras que serán más oscuros aquellos que se perciben como pequeños. Cabe resaltar que la imagen simula una representación de escala continua, por lo tanto solo serán representados como objetos negros, aquellos puntos aislados como voxels debidos al ruido.

Los datos obtenidas las imágenes caracterizadas, fueron dispuestas en un arreglo matricial para cada paciente, donde las filas de este arreglo representaban cada vóxel en la imagen y las columnas las caracterizaciones mediante las FTs. Dando como resultado una matriz de $f \times c \times s$ observaciones (f son los cortes sagitales, c los cortes axiales y s los cortes coronales de la MRI) y $p = 10$ columnas que son cada una de las FT calculadas (incluida la intensidad original). Sobre esta matriz se centralizó con la media de cada columna y posteriormente se estandarizó por desviación estándar (*Z-Score*). Los resultados obtenidos en posteriores estados del montaje experimental se realizaron con esta matriz centralizada.

6.2. Análisis de relevancia

Los resultados presentados en esta sección muestran la comparación entre los métodos basados en variabilidad (mencionados anteriormente) contra aquellos métodos de búsqueda secuencial. Los métodos de selección empleados aquí son PCA, RELIEF, PL, selección hacia adelante, eliminación hacia atrás, búsqueda secuencial flotante. Estas técnicas fueron empleadas inicialmente sobre las bases de datos más comunes mencionadas en el estado del arte (Tabla 5.2) para comprobar la adecuada capacidad de reducción con mínima redundancia y máxima relevancia en las técnicas propuestas.

Base de datos\Método	FS	BE	BEFS	PCA	Relief	PL	JMI _S	JMI _{$\alpha=1$}	JMI _{$\alpha=2$}	mRMR _S	mRMR _{$\alpha=1$}	mRMR _{$\alpha=2$}
Cardiac Arrhythmia	38.4	44.3	44.3	35.1	39.5	20.9	22.3	30.6	31.7	31.2	33.8	34.2
Australian Credit Approval	64.2	57.1	64.2	35.7	42.8	35.7	50.0	57.1	50.0	42.8	50.0	57.1
Breast Tissue	77.7	66.6	66.6	44.4	22.2	66.6	77.7	77.7	88.8	66.6	77.7	88.8
Dermatology	17.6	32.3	35.3	58.8	52.9	64.7	64.7	61.7	64.7	61.7	58.8	67.6
German Credit	87.5	83.0	87.5	50.0	50.0	60.0	58.3	45.8	54.1	50.0	45.8	58.3
Glass Identification	60.0	40.0	40.0	50.0	50.0	60.0	40.0	50.0	60.0	40.0	30.0	50.0
Iris Plants	75.0	75.0	75.0	50.0	50.0	25.0	50.0	50.0	50.0	50.0	50.0	50.0
LIBRAS movement	68.8	76.6	70.0	41.1	41.1	40.0	41.1	46.6	54.4	51.1	47.7	53.3
Indian Liver Patient	50.0	50.0	60.0	40.0	40.0	30.0	40.0	50.0	40.0	40.0	40.0	30.0
Lung Cancer	60.7	42.8	53.6	48.2	50.0	57.1	66.1	53.6	53.6	48.2	51.7	62.5
Parkinson Disease	78.3	86.9	82.6	56.5	69.5	56.5	65.2	60.8	56.5	43.8	17.8	39.1
Pima Indians Diabetes	75.0	62.5	62.5	50.0	37.5	50.0	50.0	37.5	50.0	50.0	37.5	25.0
Sonar, Mines vs. Rocks	38.3	71.6	60.0	50.0	45.0	46.6	56.0	53.0	51.6	48.3	46.0	30.0
Wine Recognition	58.3	75.0	25.0	50.0	30.0	41.6	30.0	30.0	50.0	16.6	25.0	16.6
Protein Localization Sites (Yeast)	12.5	0	12.5	12.5	0	12.5	37.5	12.5	25.0	12.5	12.5	25.0

Tabla 6.1: Resultados de la selección sobre las bases de datos del estado del arte, en términos de porcentaje de reducción.

En la Tabla 6.1 se puede encontrar una comparación directa sobre los métodos convencionales de análisis de relevancia y los métodos propuestos en este documento, estos resultados son porcentajes de reducción (porcentaje de características removidas). Se debe tener en cuenta que los resultados mostrados en los métodos JMI_S y mRMR_S fueron implementados como los describe el estado del arte en su versión estandarizada, sin embargo, tanto en estas versiones como en las propuestas en este trabajo (JMI _{$\alpha=1$} , JMI _{$\alpha=2$} , mRMR _{$\alpha=1$} y JMI _{$\alpha=2$}) fueron optimizadas utilizando el algoritmo COBRA (Sección 4.4.1).

Además, la sintonización del parámetro θ en la ecuación (5.2), que controla la interacción entre la relevancia y la redundancia, afecta directamente la selección de características. La base de datos Iris fue utilizada para demostrar la influencia de este parámetro en la selección. A esta base de datos se le aplicó *Z-Score* y se añadieron 20 características adicionales, las primeras 8 fueron señales de ruido de una distribución normal de μ 0 y σ 1, las siguientes 4 características fueron las originales trasladadas y escaladas por una constante para variar su μ y σ , las 4 finales fueron las cuatro características originales y por último las 4 siguientes fueran las originales pero contaminadas con ruido media 0 y desviación estándar de 0.3. Las

primeras 8 características se añadieron con el motivo de generar en el análisis información no relevante, después 4 siguientes se utilizan para añadir al análisis redundancia y evidenciar el comportamiento de los métodos ante la varianza. Las cuatro características contaminadas tienen como finalidad el generar 4 características compartan un porcentaje de redundancia con las originales. Y por último las 4 finales añaden 100 % de redundancia. La Figura 6.3 muestra el comportamiento inicial del parámetro cuando se añaden iterativamente cada característica de las descritas anteriormente.

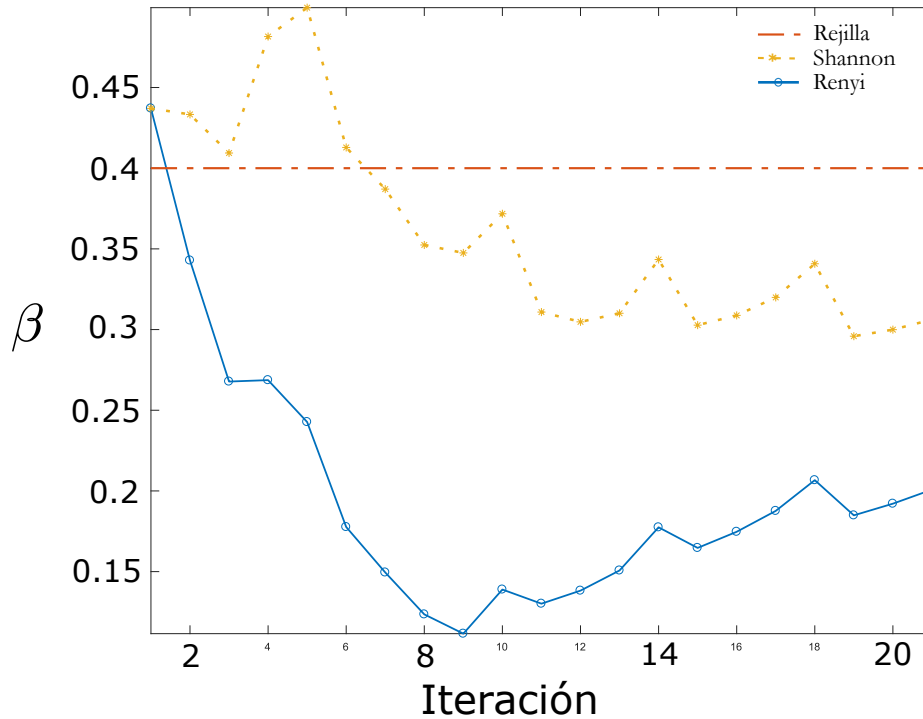


Figura 6.3: Comparación del comportamiento del parámetro.

El parámetro de la línea azul fue fijado a 0.4 después de encontrar su valor de máximo rendimiento en una rejilla de 0.1 a 0.5, este procedimiento se realizó como establece el estado del arte. En cuanto a los dos siguientes ambos fueron calculados usando la ecuación (5.2). Ambos varían en la forma como se calcula la entropía. Para la aproximación de Shannon se efectuaron histogramas con el fin de calcular las densidades de probabilidad, y para la aproximación de Renyi, se utilizó la teoría expuesta en la Sección 4.3.1, que aproxima las entropías mediante el cálculo de los kernel. Las Figuras 6.6, 6.4 y 6.5 muestran los resultados de selección de características utilizando los valores del parámetro calculados anteriormente, donde la configuración de la base de datos controlada es la siguiente:

Por color: según el color del símbolo este se debe a uno de los tipos de características presente en la base de datos de Iris controlada:

- Magenta: características originales.
- Verde: características ruidosas.
- Rojo: características escaladas.
- Azul: características escaladas y trasladadas.

Tipo del símbolo: cada una de las características de la base de datos Iris controlada, está representada por un símbolo:

- Asterisco (*), es la longitud del sépalo.
- Círculo (o), es el ancho del sépalo.
- Cruz (+), es la longitud del pétalo.
- Diamante (\diamond), es el ancho del pétalo.
- Punto (\cdot), es ruido.

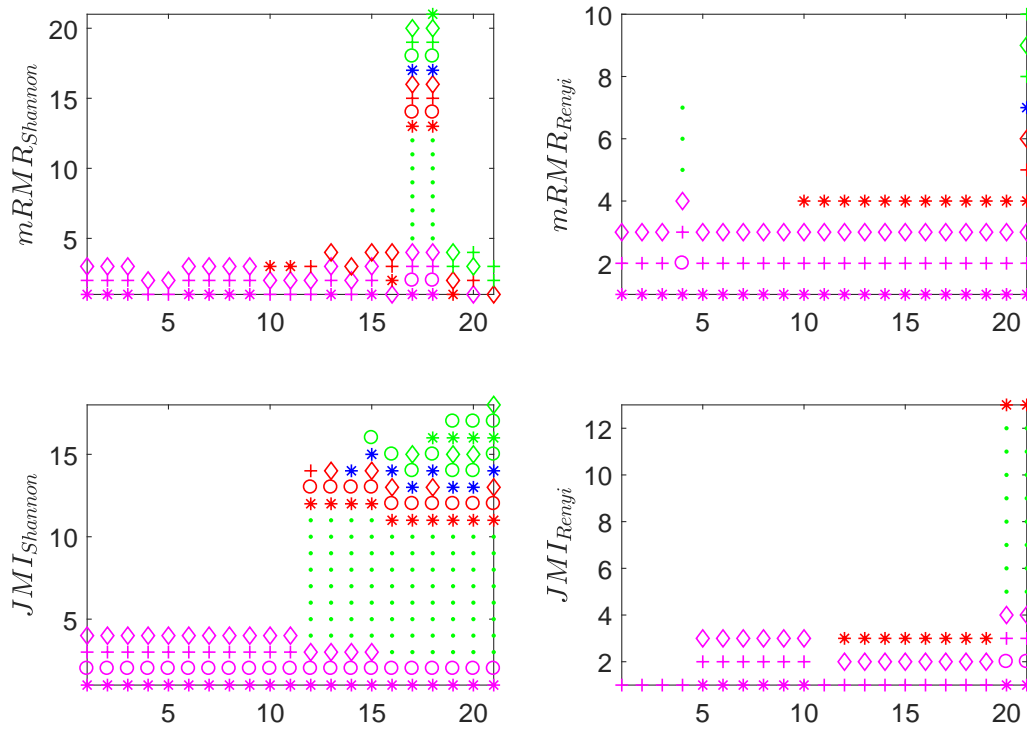


Figura 6.4: Selección de los criterios calculando el parámetro mediante histogramas

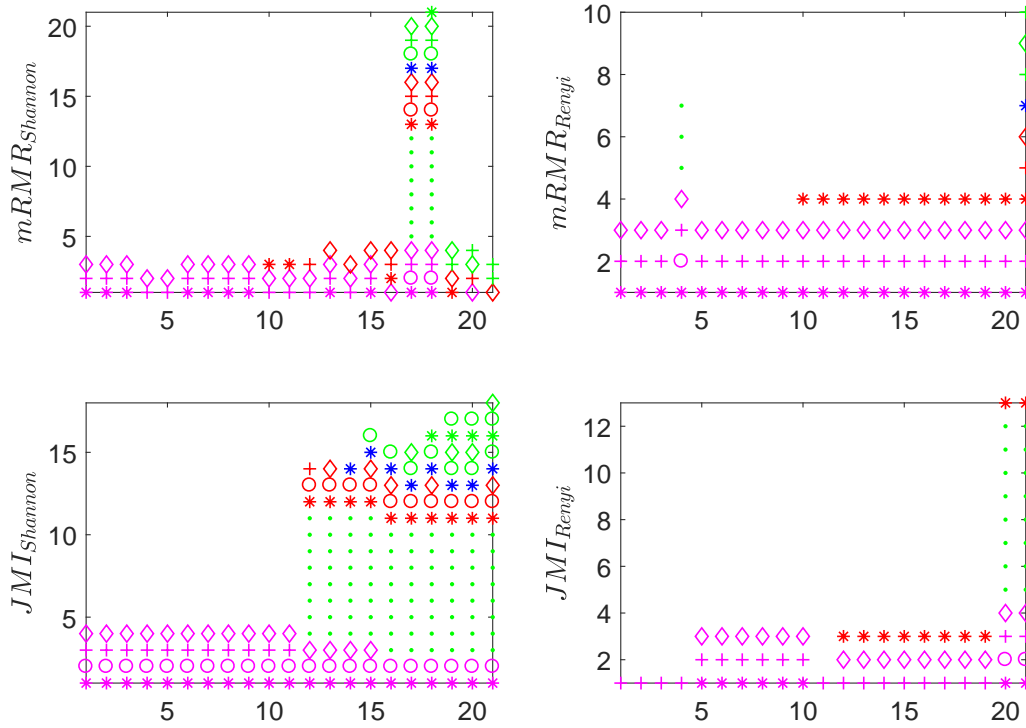


Figura 6.5: Selección de los criterios calculando el parámetro mediante entropía de Renyi

Aunque son evidentes las diferencias entre las aproximaciones respecto a la sintonización del parámetro de control entre relevancia y redundancia, se puede también resaltar que debido a la naturaleza de la selección, es complicado eliminar por completo la redundancia y seleccionar solo lo que es relevante para el experimento.

En ninguna de las versiones de PCA, LP o Relief, se hizo sintonización de parámetros, pues se implementaron según las versiones de estos algoritmos que referencia el estado del arte.

Aunque el porcentaje de reducción de los métodos de información mutua basados en entropía de Renyi con $\alpha = 2$ propuestos aquí, presentan los mejores los potenciales de reducción, es importante resaltar que existen casos específicos en los cuales los métodos de búsqueda secuencial logran porcentajes de reducción mucho mayores y con mejores resultados de clasificación.

En cuento a la selección sobre la base de datos MRI de la Tabla 5.1, la Tabla 6.2 muestra que

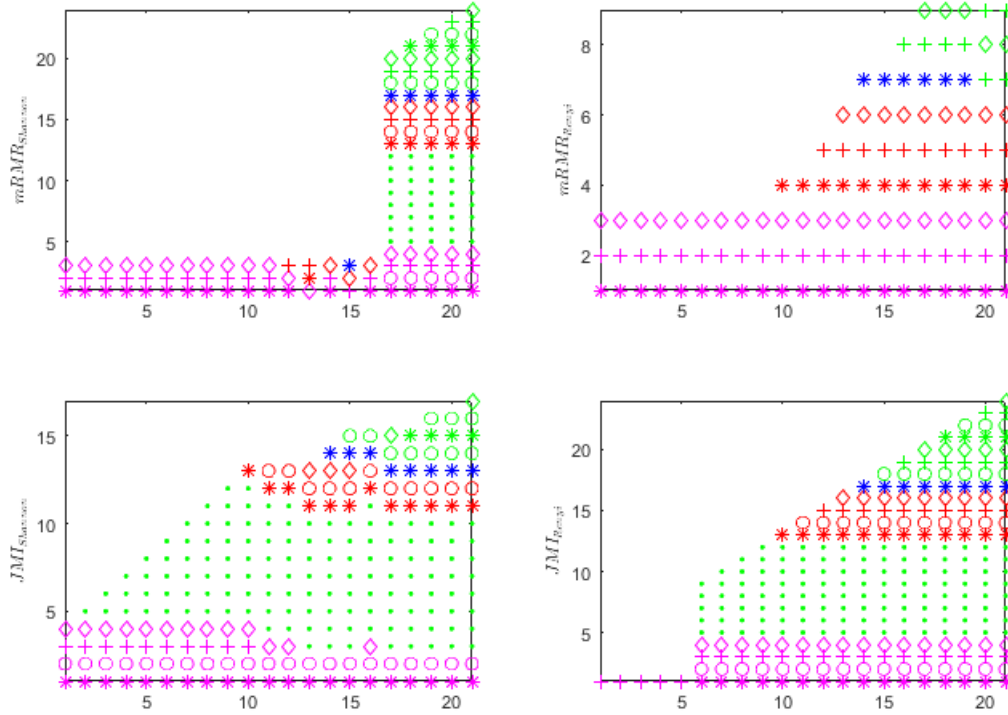


Figura 6.6: Selección de los criterios utilizando el parámetro fijo

nuestros métodos propuestos de selección pueden hacer una reducción de características más potente que los métodos clásicos, pasando en el caso del primer paciente de aproximadamente 44 millones de datos por 10 características, a alrededor de 33 % de datos.

Paciente	FS	BE	BEFS	PCA	Relief	LP	JMI _S	JMI _{α=1}	JMI _{α=2}	mRMR _S	mRMR _{α=1}	mRMR _{α=2}
1	36.23	21.98	41.56	11.37	23.76	0.87	49.76	48.01	78.02	47.81	50.23	69.03
2	41.08	34.01	62.24	26.52	28.9	34.82	50.19	51.23	67.16	55.2	52.49	64.11
3	28.11	13.24	38.67	7.76	11.81	25.13	40.65	42.3	49.89	41.31	40.17	50.27

Tabla 6.2: Porcentaje de reducción de datos sobre los pacientes de DCF para cada método de análisis de relevancia.

6.3. Clasificación

Con el objetivo de evaluar la eficacia en términos de clasificación presentada en la sección 6.2, la Tablas 6.3 a 6.7 muestran la clasificación de los espacios generados para las bases de datos sin desbalance o con un orden de desbalance no mayor de 1:3 según la tabla 5.2. Los clasificadores utilizados para estos datos fueron los mencionados en la sección 5.4, y cabe resaltar que ninguno de los parámetros de los clasificadores fueron sintonizados. Es importante aclarar que la última fila de las tablas antes mencionadas presentan los resultados de clasificación sin aplicar la etapa de selección de características, estos resultados se representan con las letras «NFS».

Método	$k - NN(5)$	LDA	SVM	DT
FS	45.40±14.20	24.00±2.57	24.17±2.59	35.17±3.60
BE	39.10±2.76	31.30±2.49	30.70±2.21	36.53±3.84
BEFS	46.51±11.15	31.20±8.60	28.46±4.13	40.87±9.68
PCA	70.83±2.07	76.13±2.31	75.97±2.34	69.80±2.70
Relief	69.67±2.30	74.80±1.21	73.60±2.02	69.87±2.23
LP	75.13±1.87	75.13±1.87	75.13±1.87	75.13±1.87
JMI _S	70.83±2.07	76.13±2.31	75.90±2.46	69.63±2.45
JMI _{α=1.001}	69.23±1.78	74.47±3.24	74.87±2.99	67.63±3.48
JMI _{α=2}	77.28±2.81	78.72±2.11	81.15±4.19	79.33±4.12
mRMR _S	71.33±2.83	74.97±2.96	75.37±3.05	71.20±2.31
mRMR _{α=1.001}	69.90±1.52	74.67±2.95	74.07±3.19	69.17±3.55
mRMR _{α=2}	76.15±0.46	77.02±1.64	78.39±4.68	78.16±2.24
NSF	75.13±1.48	76.13±2.31	72.64±3.10	76.86±2.85

Tabla 6.3: Germa, No. clases 2, No. muestras 1000, No. características 24

Método	$k - NN(5)$	LDA	SVM	DT
FS	48.67±4.85	48.67±8.28	47.56±8.26	49.33±6.77
BE	22.00±2.86	15.78±8.14	16.89±6.96	18.44±10.21
BEFS	39.59±6.53	35.77±14.28	33.25±9.34	38.91±14.45
PCA	95.11±2.04	98.22±2.04	96.67±2.16	94.22±2.39
Relief	82.67±7.61	89.11±7.66	86.44±8.85	82.00±11.16
PL	96.22±1.83	98.22±1.25	98.16±3.64	95.64±3.44
JMI _S	95.33±2.21	98.00±2.21	96.22±2.11	94.67±2.81
JMI _{$\alpha=1.001$}	95.78±1.64	95.56±1.48	95.78±1.95	93.56±1.64
JMI _{$\alpha=2$}	96.36±2.43	97.23±0.46	98.36±3.14	96.16±3.01
mRMR _S	96.89± 1.55	96.22±1.83	96.00±2.30	94.67±2.81
mRMR _{$\alpha=1.001$}	96.22±1.50	98.22±2.04	96.44±1.87	94.22±2.39
mMMR _{$\alpha=2$}	95.63±2.31	98.23±3.66	99.01±5.42	97.31±2.64
NSF	96.22±2.04	98.22±1.25	98.16±3.64	95.64±3.44

Tabla 6.4: Iris, No. clases 3, No. muestras 150, No. características 4

Método	$k - NN(5)$	LDA	SVM	DT
FS	93.89±1.32	92.69±1.19	91.76±1.82	92.50±1.02
BE	31.11±3.06	42.59±2.58	22.04±3.90	47.31±4.89
BEFS	66.76±4.86	71.19±7.96	57.93±4.59	74.93±8.91
PCA	69.26±4.27	60.46±2.51	76.57±3.73	59.54±6.73
Relief	69.26±4.04	62.69±2.80	77.31±3.24	62.22±7.28
PL	68.61±2.81	68.61±2.81	68.61±2.81	68.61±2.81
JMI _S	69.26±4.27	60.46±2.51	76.57±3.73	60.09±6.30
JMI _{$\alpha=1.001$}	69.07±4.30	60.09±2.60	76.39±3.44	60.09±6.30
mRMR _S	69.26±4.27	60.46±2.51	76.57±3.73	60.09±6.30
mRMR _{$\alpha=1.001$}	66.85±5.67	59.81±4.19	76.57±2.58	58.89±5.78
NSF	60.46±2.51	60.46±2.51	60.46±2.51	60.46±2.51

Tabla 6.5: Libras, No. clases 15, No. muestras 360, No. características 90

Método	$k - NN(5)$	LDA	SVM	DT
FS	39.25±2.23	29.71±1.09	28.74±0.00	39.66±1.71
BE	39.43±1.41	30.98±1.45	28.74±0.00	39.94±1.70
BEFS	43.60±4.49	33.89±7.33	29.76±1.73	44.82±7.66
PCA	66.95±4.99	69.54±2.85	70.69±1.82	65.63±3.71
Relief	66.95±4.99	69.54±2.85	70.69±1.82	65.69±4.09
PL	71.15±1.08	71.15±1.08	71.15±1.08	71.15±1.08
JMI _S	67.18±4.46	69.66±2.47	71.26±0.00	63.23±3.29
JMI _{α=1.001}	66.95±4.86	69.83±2.60	70.69±1.82	65.75±3.64
mRMR _S	66.95±4.99	69.54±2.85	70.69±1.82	63.80±3.57
mRMR _{α=1.001}	67.07±4.85	69.83±2.81	70.69±1.82	66.32±2.79
NSF	69.54±2.85	69.54±2.85	69.54±2.85	69.54±2.85

Tabla 6.6: Liver, No. clases 2, No. muestras 583, No. Características 9

Método	$k - NN(5)$	LDA	SVM	DT
FS	36.38±3.76	31.72±3.56	26.38±3.36	40.34±2.95
BE	16.03±4.15	21.21±5.86	19.66±4.75	21.38±6.41
BEFS	30.47±6.63	30.01±10.78	24.05±5.78	35.89±10.64
PCA	91.03±4.13	85.69±2.82	86.55±2.67	82.07±6.81
Relief	80.17±6.86	76.21±5.32	76.03±4.97	75.52±6.49
PL	91.55±3.49	84.48±3.35	87.07±3.07	85.00±2.00
JMI _S	92.59±3.73	84.31±3.94	87.59±2.67	81.55±4.88
JMI _{α=1.001}	91.03±4.13	85.69±2.82	86.55±2.67	82.76±6.75
mRMR _S	88.97±3.83	83.79±3.65	84.14±3.13	85.34±5.41
mRMR _{α=1.001}	84.66±2.87	84.31±3.68	85.17±2.02	82.59±4.02
NSF	91.03±4.13	85.69±2.82	86.55±2.67	82.76±6.75

Tabla 6.7: Parkinson, No. clases 2, No. muestras 195, No. características 22

La Tabla 6.8 muestra los resultados de selección para las bases de datos que exhiben un alto desbalance de datos. Tanto en los datos desbalanceados como en los balanceados, se puede encontrar que los métodos basados en información mutua obtienen desempeños mayores de clasificación, especialmente el método de información mutua conjunta (JMI), que en muchas de las pruebas arroja rendimientos altos con muy pocas características según la Tabla 6.1.

Clasificación con ensamble (DT+RUSBoost)							
Base desbalanceada	FS	BE	BEFS	PCA	Relief	PL	JMI _S
Cardiac Arrhythmia	68,16 ± 3,25	73,13 ± 5,26	73,13 ± 5,26	69,94 ± 0,73	70,31 ± 2,25	76,72 ± 4,31	72,09 ± 1,94
Australian Credit Approval	49,74 ± 7,00	60,99 ± 3,87	76,14 ± 6,48	85,76 ± 3,42	86,79 ± 3,28	91,03 ± 1,34	88,64 ± 3,36
Breast Tissue	96,97 ± 2,54	81,00 ± 1,05	95,43 ± 3,59	90,48 ± 5,94	87,91 ± 3,46	94,39 ± 3,59	91,63 ± 2,94
Dermatology	15,96 ± 9,51	62,20 ± 9,15	66,11 ± 7,94	94,59 ± 1,70	86,61 ± 5,18	95,96 ± 1,79	95,50 ± 1,64
Glass Identification	68,75 ± 2,44	88,03 ± 4,30	88,03 ± 4,30	80,47 ± 5,91	79,38 ± 6,80	77,53 ± 0,65	76,81 ± 2,23
Indian Liver Patient	60,91 ± 7,97	97,32 ± 2,73	97,32 ± 2,73	90,36 ± 3,08	97,32 ± 2,31	92,36 ± 3,54	99,01 ± 0,19
Lung Cancer	40,63 ± 17,6	95,16 ± 8,09	97,23 ± 10,6	80,44 ± 11,0	80,43 ± 13,2	85,76 ± 10,6	94,49 ± 9,78
Pima Indians Diabetes	45,13 ± 2,24	60,61 ± 12,42	60,61 ± 12,42	96,79 ± 3,31	91,16 ± 2,30	96,97 ± 2,34	97,31 ± 0,14
Sonar, Mines vs. Rocks	81,02 ± 2,13	60,91 ± 4,43	67,34 ± 1,47	79,09 ± 2,53	73,87 ± 5,26	74,43 ± 3,15	75,49 ± 1,38
Wine Recognition	60,57 ± 6,62	16,42 ± 4,45	89,09 ± 3,71	90,19 ± 3,65	90,25 ± 4,18	91,13 ± 3,67	90,36 ± 4,20
Protein Localization Sites (Yeast)	74,33 ± 1,60	79,92 ± 4,11	88,15 ± 6,88	78,92 ± 1,27	77,63 ± 5,21	67,72 ± 5,36	78,57 ± 4,81

					DT	
					NSF	NSF
Cardiac Arrhythmia	JMI _{α=1,001}	JMI _{α=2}	mRMR _S	mRMR _{α=1,001}	mRMR _{α=2}	51,13 ± 1,67
Australian Credit Approval	71,84 ± 3,26	75,34 ± 2,31	69,23 ± 1,97	69,29 ± 3,12	82,69 ± 0,97	85,51 ± 2,86
Breast Tissue	88,94 ± 3,61	91,13 ± 0,69	89,14 ± 2,13	88,23 ± 2,16	87,36 ± 1,30	60,97 ± 6,53
Dermatology	89,09 ± 4,37	97,73 ± 1,91	94,29 ± 3,22	93,73 ± 1,57	97,29 ± 2,05	30,49 ± 1,73
Glass Identification	94,86 ± 1,51	95,63 ± 3,01	94,86 ± 2,21	91,14 ± 18,22	95,91 ± 1,76	61,88 ± 3,31
Indian Liver Patient	75 ± 3,26	87,29 ± 4,13	73,63 ± 2,89	76,64 ± 4,62	70,86 ± 3,01	69,54 ± 2,85
Lung Cancer	98,36 ± 1,02	99,49 ± 0,04	95,94 ± 3,63	96,42 ± 0,26	99,01 ± 0,03	46,67 ± 17,2
Pima Indians Diabetes	94,13 ± 9,06	95,36 ± 5,31	92,09 ± 8,68	90,76 ± 11,6	93,12 ± 9,49	71,74 ± 2,97
Sonar, Mines vs. Rocks	97,19 ± 1,82	99,46 ± 0,10	97,31 ± 0,14	98,29 ± 0,61	1,00 ± 0,00	73,87 ± 5,26
Wine Recognition	77,21 ± 2,13	80,94 ± 1,63	75,69 ± 4,84	76,37 ± 3,92	74,63 ± 4,71	88,57 ± 3,67
Protein Localization Sites (Yeast)	90,36 ± 4,20	90,36 ± 4,20	91,03 ± 1,86	90,91 ± 6,11	92,01 ± 5,12	54,88 ± 2,52
	77,93 ± 1,49	89,43 ± 2,49	78,39 ± 2,63	81,50 ± 4,06	76,54 ± 1,87	79,50 ± 3,71

Tabla 6.8: Clasificación de datos desbalanceados en términos de la media geométrica.

Capítulo 7

Conclusiones y trabajo futuro

7.1. Conclusiones

1. Se desarrolló una metodología de segmentación automática de corteza cerebral en MRI usando características volumétricas por renderizado de FT, basado en información mutua que puede apoyar el diagnóstico clínico neurológico de displasias corticales focales.
2. Se comprobó que el análisis de relevancia por información mutua tiene una potencia de reducción de características mayor que los métodos de selección por puntuación, debido a que muchos de los métodos convencionales de puntuación tienen su fundamento en la variabilidad de los datos y en su estructura.
3. Se determinó que el uso de los criterios basados en información mutua alcanzan porcentajes de reducción mayor que los criterios de optimización soportados en algoritmos de búsqueda secuencial, haciendo a estos criterios de información mutua medidas adecuadas para el análisis de relevancia cuando se usan con estos algoritmos de búsqueda secuencial.
4. El estrategia COBRA presentó mejores resultados de optimización que los algoritmos de búsqueda secuencial, mostrando que la programación semidefinida alcanza soluciones estables y apropiadas para el problema de selección de características.
5. Se identificaron las anormalidades en los pacientes presentados en este trabajo, alcanzando una media geométrica promedio por encima del 89 %.
6. Los resultados de clasificación demuestran que tanto para datos balanceados como para los no balanceados, la implementación de una etapa de análisis de relevancia aporta mejoras de rendimiento sustanciales dependiendo del método de selección utilizado.

Esto se puede ver reflejado en la Tabla 6.8 que muestra que para un clasificador simple como DT, la etapa de selección ayuda a mejorar su precisión.

7. Se encontró que la clasificación de datos desbalanceados y de gran volumen se ve potenciada por el uso de ensambles. Especialmente el algoritmo RUSBoost demuestra ser una herramienta de alto impacto en términos de clasificación grandes cantidades de datos.

7.2. Trabajo futuro

Como trabajo futuro, se propone la sintonización de parámetros asociados al algoritmo COBRA que podrían mejorar la optimización de la selección de características. También se propone la optimización del cálculo de información mutua basada en Renyi, debido a que aunque el cálculo de las entropías ofrecieron resultados más eficientes en términos de reducción que los histogramas, el cálculo de los kernel para bases de datos de muchas muestras es un proceso que demanda mucha capacidad de almacenamiento.

Bibliografía

- [1] W. H. Organization *et al.*, “Global burden of epilepsy and the need for coordinated action at the country level to address its health, social and public knowledge implications [internet]. [cited 2018 may 2].”
- [2] J. A. Delgado-De Bedout, J. G. Castrillón-Guzmán, S. J. Rascovsky-Ramirez, J. M. Vélez-Arango, and V. D. Calvo-Betancur, “Detección de displasias corticales asistida mediante métodos semiautomáticos de espesor cortical,” *Salud Uninorte*, vol. 31, no. 3, 2015.
- [3] J. Ashburner, “Computational anatomy with the spm software,” *Magnetic resonance imaging*, vol. 27, no. 8, pp. 1163–1174, 2009.
- [4] M. M. Eapena, M. S. J. A. Ancelita, and G. Geetha, “Segmentation of tumors from ultrasound images with paorgb,” *Procedia Computer Science*, vol. 50, pp. 663–668, 2015.
- [5] B. Foster, U. Bagci, A. Mansoor, Z. Xu, and D. J. Mollura, “A review on segmentation of positron emission tomography images,” *Computers in biology and medicine*, vol. 50, pp. 76–96, 2014.
- [6] U. Bagci, J. K. Udupa, N. Mendhiratta, B. Foster, Z. Xu, J. Yao, X. Chen, and D. J. Mollura, “Joint segmentation of anatomical and functional images: Applications in quantification of lesions from pet, pet-ct, mri-pet, and mri-pet-ct images,” *Medical image analysis*, vol. 17, no. 8, pp. 929–945, 2013.
- [7] C. Chen, W. Xie, J. Franke, P. Grutzner, L.-P. Nolte, and G. Zheng, “Automatic x-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements,” *Medical image analysis*, vol. 18, no. 3, pp. 487–499, 2014.
- [8] D. Taylor, M. Falconer, C. Bruton, and J. Corsellis, “Focal dysplasia of the cerebral cortex in epilepsy,” *Journal of Neurology, Neurosurgery & Psychiatry*, vol. 34, no. 4, pp. 369–387, 1971.

- [9] J. N. Gaitanis and J. Donahue, “Focal cortical dysplasia,” *Pediatric neurology*, vol. 49, no. 2, pp. 79–87, 2013.
- [10] Y. Artan, A. Oto, and I. S. Yetik, “Cross-device automated prostate cancer localization with multiparametric mri,” *IEEE Transactions on Image Processing*, vol. 22, no. 12, pp. 5385–5394, 2013.
- [11] L. Yang, O. Tuzel, P. Meer, and D. J. Foran, “Automatic image analysis of histopathology specimens using concave vertex graph,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2008, pp. 833–841.
- [12] H. Pfister, B. Lorensen, C. Bajaj, G. Kindlmann, W. Schroeder, L. S. Avila, K. Raghu, R. Machiraju, and J. Lee, “The transfer function bake-off,” *IEEE Computer Graphics and Applications*, vol. 21, no. 3, pp. 16–22, 2001.
- [13] C. Correa and K.-L. Ma, “Size-based transfer functions: A new volume exploration technique,” *IEEE transactions on visualization and computer graphics*, vol. 14, no. 6, pp. 1380–1387, 2008.
- [14] J. Kniss, G. Kindlmann, and C. Hansen, “Interactive volume rendering using multi-dimensional transfer functions and direct manipulation widgets,” in *Proceedings of the Conference on Visualization’01*. IEEE Computer Society, 2001, pp. 255–262.
- [15] P. Ljung, J. Krüger, E. Groller, M. Hadwiger, C. D. Hansen, and A. Ynnerman, “State of the art in transfer functions for direct volume rendering,” in *Computer Graphics Forum*, vol. 35, no. 3. Wiley Online Library, 2016, pp. 669–691.
- [16] K. Engel, M. Hadwiger, J. Kniss, C. Rezk-Salama, and D. Weiskopf, *Real-time volume graphics*. CRC Press, 2006.
- [17] T. Pfaffelmoser, M. Reitering, and R. Westermann, “Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields,” in *Computer Graphics Forum*, vol. 30, no. 3. Wiley Online Library, 2011, pp. 951–960.
- [18] Y. Wang, J. Zhang, D. J. Lehmann, H. Theisel, and X. Chi, “Automating transfer function design with valley cell-based clustering of 2d density plots,” in *Computer Graphics Forum*, vol. 31, no. 3pt4. Wiley Online Library, 2012, pp. 1295–1304.
- [19] I. Fujishiro, T. Azuma, and Y. Takeshima, “Automating transfer function design for comprehensible volume rendering based on 3d field topology analysis,” in *Visualization’99. Proceedings*. IEEE, 1999, pp. 467–563.
- [20] I. Fujishiro, Y. Takeshima, T. Azuma, and S. Takahashi, “Volume data mining using 3d field topology analysis,” *IEEE Computer Graphics and Applications*, vol. 20, no. 5, pp. 46–51, 2000.

- [21] L. Wang and A. Kaufman, "Importance driven automatic color design for direct volume rendering," in *Computer graphics forum*, vol. 31, no. 3pt4. Wiley Online Library, 2012, pp. 1305–1314.
- [22] R. Bramon, M. Ruiz, A. Bardera, I. Boada, M. Feixas, and M. Sbert, "Information theory-based automatic multimodal transfer function design," *IEEE journal of biomedical and health informatics*, vol. 17, no. 4, pp. 870–880, 2013.
- [23] T. M. Cover, "The best two independent measurements are not the two best," *IEEE Transactions on Systems, Man, and Cybernetics*, no. 1, pp. 116–117, 1974.
- [24] T. M. Cover and J. A. Thomas, *Elements of information theory*. John Wiley & Sons, 2012.
- [25] A. K. Jain, R. P. Duin, and J. Mao, "Statistical pattern recognition: A review," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 1, pp. 4–37, 2000.
- [26] A. R. Webb, *Statistical pattern recognition*. John Wiley & Sons, 2003.
- [27] M. J. Mendenhall and E. Merényi, "Relevance-based feature extraction for hyperspectral images," *IEEE Transactions on Neural Networks*, vol. 19, no. 4, pp. 658–672, 2008.
- [28] M. Imani and H. Ghassemian, "Feature extraction using attraction points for classification of hyperspectral images in a small sample size situation," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 11, pp. 1986–1990, 2014.
- [29] S. Bruckner, S. Grimm, A. Kanitsar, and M. E. Groller, "Illustrative context-preserving exploration of volume data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 6, pp. 1559–1569, 2006.
- [30] E. K. Fishman, D. R. Ney, D. G. Heath, F. M. Corl, K. M. Horton, and P. T. Johnson, "Volume rendering versus maximum intensity projection in ct angiography: what works best, when, and why," *Radiographics*, vol. 26, no. 3, pp. 905–922, 2006.
- [31] M. Habbecke and L. Kobbelt, "Iterative multi-view plane fitting," in *Int. Fall Workshop of Vision, Modeling, and Visualization*, 2006, pp. 73–80.
- [32] A. Opitz, *Classification and visualization of volume data using clustering*. na, 2009.
- [33] B. Girod, A. Stenger, and R. Rabenstein, *Signals and systems*. Wiley, 2001.
- [34] C. S. Williams and O. A. Becklund, *Introduction to the optical transfer function*. SPIE Press, 1989, vol. 112.

- [35] G. Kindlmann and J. W. Durkin, “Semi-automatic generation of transfer functions for direct volume rendering,” in *Proceedings of the 1998 IEEE symposium on Volume visualization*. ACM, 1998, pp. 79–86.
- [36] G. A. Korn and T. M. Korn, *Mathematical handbook for scientists and engineers: Definitions, theorems, and formulas for reference and review*. Courier Corporation, 2000.
- [37] J. E. Marsden and A. Tromba, *Vector calculus*. Macmillan, 2003.
- [38] J. R. Dormand and P. J. Prince, “A family of embedded runge-kutta formulae,” *Journal of computational and applied mathematics*, vol. 6, no. 1, pp. 19–26, 1980.
- [39] M. Levoy, “Direct visualization of surfaces from computed tomography data,” in *Medical Imaging II*, vol. 914. International Society for Optics and Photonics, 1988, pp. 828–842.
- [40] G. Kindlmann, R. Whitaker, T. Tasdizen, and T. Moller, “Curvature-based transfer functions for direct volume rendering: Methods and applications,” in *Visualization, 2003. VIS 2003. IEEE*. IEEE, 2003, pp. 513–520.
- [41] O. Monga, S. Benayoun, and O. D. Faugeras, “From partial derivatives of 3-d density images to ridge lines,” in *Computer Vision and Pattern Recognition, 1992. Proceedings CVPR’92., 1992 IEEE Computer Society Conference on*. IEEE, 1992, pp. 354–359.
- [42] Z. Nagy, J. Schneider, and R. Westermann, “Interactive volume illustration.” in *VMV*, 2002, pp. 497–504.
- [43] V. Interrante, “Illustrating surface shape in volume data via principal direction-driven 3d line integral convolution,” in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 1997, pp. 109–116.
- [44] J. J. Caban and P. Rheingans, “Texture-based transfer functions for direct volume rendering,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1364–1371, 2008.
- [45] M. Haidacher, D. Patel, S. Bruckner, A. Kanitsar, and M. E. Gröller, “Volume visualization based on statistical transfer-function spaces,” in *Visualization Symposium (Pacific Vis), 2010 IEEE Pacific*. IEEE, 2010, pp. 17–24.
- [46] F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll, “Image selective smoothing and edge detection by nonlinear diffusion,” *SIAM Journal on Numerical analysis*, vol. 29, no. 1, pp. 182–193, 1992.
- [47] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” *IEEE Transactions on pattern analysis and machine intelligence*, vol. 12, no. 7, pp. 629–639, 1990.

- [48] D. Mandal, A. Chatterjee, and M. Maitra, “Robust medical image segmentation using particle swarm optimization aided level set based global fitting energy active contour approach,” *Engineering Applications of Artificial Intelligence*, vol. 35, pp. 199–214, 2014.
- [49] A. M. Álvarez-Meza, L. F. Velásquez-Martínez, and G. Castellanos-Dominguez, “Time-series discrimination using feature relevance analysis in motor imagery classification,” *Neurocomputing*, vol. 151, pp. 122–129, 2015.
- [50] S. Bhattacharyya, A. Sengupta, T. Chakraborti, A. Konar, and D. Tibarewala, “Automatic feature selection of motor imagery eeg signals using differential evolution and learning automata,” *Medical & biological engineering & computing*, vol. 52, no. 2, pp. 131–139, 2014.
- [51] X. Wang and K. K. Paliwal, “Feature extraction and dimensionality reduction algorithms and their applications in vowel recognition,” *Pattern recognition*, vol. 36, no. 10, pp. 2429–2439, 2003.
- [52] G. Daza-Santacoloma, J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruíz, and G. Castellanos-Domínguez, “Dynamic feature extraction: an application to voice pathology detection,” *Intelligent Automation & Soft Computing*, vol. 15, no. 4, pp. 667–682, 2009.
- [53] J. R. Orozco-Arroyave, S. Murillo-Rendón, A. M. Álvarez-Meza, J. D. Arias-Londoño, E. Delgado-Trejos, J. Vargas-Bonilla, and C. G. Castellanos-Domínguez, “Automatic selection of acoustic and non-linear dynamic features in voice signals for hypernasality detection,” in *Twelfth Annual Conference of the International Speech Communication Association*, 2011.
- [54] O. Shamir, “A stochastic pca and svd algorithm with an exponential convergence rate,” in *International Conference on Machine Learning*, 2015, pp. 144–152.
- [55] M. Belkin and P. Niyogi, “Laplacian eigenmaps and spectral techniques for embedding and clustering,” in *Advances in neural information processing systems*, 2002, pp. 585–591.
- [56] X. He and P. Niyogi, “Locality preserving projections,” in *Advances in neural information processing systems*, 2004, pp. 153–160.
- [57] T. Naghibi, S. Hoffmann, and B. Pfister, “A semidefinite programming based search strategy for feature selection with mutual information measure,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 8, pp. 1529–1541, 2015.
- [58] R. Battiti, “Using mutual information for selecting features in supervised neural net learning,” *IEEE Transactions on neural networks*, vol. 5, no. 4, pp. 537–550, 1994.

- [59] N. Kwak and C.-H. Choi, "Input feature selection for classification problems," *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 143–159, 2002.
- [60] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 27, no. 8, pp. 1226–1238, 2005.
- [61] H. Yang and J. Moody, "Feature selection based on joint mutual information," in *Proceedings of international ICSC symposium on advances in intelligent data analysis*. Citeseer, 1999, pp. 22–25.
- [62] W. Gao, L. Hu, P. Zhang, and J. He, "Feature selection considering the composition of feature relevancy," *Pattern Recognition Letters*, vol. 112, pp. 70–74, 2018.
- [63] L. G. S. Giraldo, M. Rao, and J. C. Principe, "Measures of entropy from data using infinitely divisible kernels," *IEEE Transactions on Information Theory*, vol. 61, no. 1, pp. 535–548, 2015.
- [64] H. He and E. A. Garcia, "Learning from imbalanced data," *IEEE Transactions on knowledge and data engineering*, vol. 21, no. 9, pp. 1263–1284, 2009.
- [65] M. Maalouf and T. B. Trafalis, "Rare events and imbalanced datasets: an overview," *International Journal of Data Mining, Modelling and Management*, vol. 3, no. 4, pp. 375–388, 2011.
- [66] C. Zhang, W. Gao, J. Song, and J. Jiang, "An imbalanced data classification algorithm of improved autoencoder neural network," in *Advanced Computational Intelligence (ICA-CI), 2016 Eighth International Conference on*. IEEE, 2016, pp. 95–99.
- [67] B. Wang and J. Pineau, "Online bagging and boosting for imbalanced data streams," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 12, pp. 3353–3366, 2016.
- [68] D. Wu, Z. Wang, Y. Chen, and H. Zhao, "Mixed-kernel based weighted extreme learning machine for inertial sensor based human activity recognition with imbalanced dataset," *Neurocomputing*, vol. 190, pp. 35–49, 2016.
- [69] C. Seiffert, T. M. Khoshgoftaar, J. Van Hulse, and A. Napolitano, "Rusboost: A hybrid approach to alleviating class imbalance," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 40, no. 1, pp. 185–197, 2009.
- [70] N. V. Chawla, A. Lazarevic, L. O. Hall, and K. W. Bowyer, "Smoteboost: Improving prediction of the minority class in boosting," in *European conference on principles of data mining and knowledge discovery*. Springer, 2003, pp. 107–119.